PRELIMINARIES TO THE PREPARATION OF A SPELL AND GRAMMAR CHECKER FOR TAMIL

Dr. S.Rajendran
Associate Professor
Tamil University
Thanjavur 613 005

1. Introduction

All the word processors available for English have a tool to check the spelling errors in the text typed by one. For example, Wordstar, MS-Word has the facility to check the spelling mistakes. MS-Word has the most advanced tool to check the spelling errors in the text. MS-Word has three ways of checking spelling:

1. Many common typing errors are corrected automatically using a tool called Auto correct. One can also customize Auto correct to include common errors.
2. One can set MS-Word's spelling and grammar feature to check spelling as you type.
3. One can run a spelling check on a whole document at any time one wants.

MS-Word can check your spelling as you type. i.e. Whenever is thinks it has found a spelling error, it will underline it with a wavy red line. To see the suggested correction, right click on the underlined word and look at the words listed at the top of the pop-up menu. Select the word that you want to use and the change will be made automatically. By default, MS-Word is set to check spelling as you type. You can turn this feature off and on by selecting options from the Tools menu.

MS-Word can check your grammar as you type. i.e. Whenever it thinks it has found a grammatical error, it will underline it with a wavy green line. To see the suggested corrections, right click on the underlined word and look at the suggestions on the pop-up menu. If you want to know what type of error the grammar checker had identified click on the grammar option at the bottom of the pop-up menu. This menu pops up and says the problem is Subject-Verb Agreement. Sometimes, you will be able to select a change to be made automatically. At other times the suggestion will just help you to think how you can change the sentence yourself.

A project entitled "Spell and grammar checker for Tamil" has been undertaken to explore the possibility of preparing a machine aid to check the spelling and grammar of Tamil. The spell and grammar checker is aimed at being designed to go through the Tamil text under consideration and to indicate the spelling and grammar mistakes. If there are any spelling or grammar mistakes in the text, it will give suggestions and/or correction in terms of options.

The spell and grammar checker needs to have at least the following impediments:

1. Preparation of a dictionary
2. Preparation of a morphological analyser
3. Preparation of a syntactic analyser
4. Building of an interface to speel-check and grammar-chek the text

It is a pan Indian feature found in many Indian languages (such as Tamil, Malayalam, Kannada, Telugu, Hindi etc.) that their verb phrase consists of different combinations of verbs culminating into compound verbs which contribute to the interpretation of tense, mood, aspect, attitude, non-attitude, voice, causation, and negative polarity in a sentence. The structure of verbal complex (which include simple and compound verbs) is unique and capturing this complexity in a machine analyzable and generatable format is a challenging job. What is workable in the forward direction may not work in the backward direction. Building such a tool that analyzes and generates the complex verbal system with the proper interpretation of meaning involves skillful linguistic analysis and computational programming. Though the formation of the verbal complex is mainly anchored on the morphological level, it is linked to syntax on the one side and semantics on the other side. The formation of the verbal complex involves arrangement of the verbal units and the interpretation of their combinatory meaning. Phonology also plays its part in the formation of verbal complex in terms of morphophonemic or sandhi rules which account for the shape changes due to inflection and juxtaposition.

2. Preparation of a dictionary

A Tamil dictionary has been prepared in the computer version. Entries are arranged in alphabetical order. The information in each individual entry is presented in the following way:

1. Head word, presented in the alphabetical order, letter by letter.
2. Grammatical category
3. Verb classes based on conjugation.
4. Noun classes based on case inflection and number inflection
5. Adjectives, adverbs and postpostions, etc. does not have classes.

3. Preparation of morphological analyser

A Tamil word can be of two types: simple and compound. A simple word consists of a root or stem together with suffixes. If we have an exhaustive lexicon showing all the word forms of all the roots along with their feature values, we do not naturally need a morphological analyser. In such a lexicon all we need to do is to look for a word in the lexicon and retrieve its feature values. The above method has several drawbacks. It is extremely wasteful of memory space of a computer. It does not show the relationship among different roots that have similar word forms. Tamil

has a rich and productive morphology and so the number of word forms might well be infinite. Most natural language processing systems use simple linguistic theories for morphological analysis. The morphological analyser takes as input a sentence, i.e. a sequence of words. For each of the words, it looks up the dictionary and retrieves information such as the root of the word, its grammatical category, person, number, gender and tense, etc. In case, a word has multiple meanings, grammatical information is retrieved for each of the meanings.

For the present work, the scheme is based on paradigms. We have provided different table of word forms covering the words in Tamil. Each word form table covers a set of roots which means that the roots follow the pattern implicit in the table for generating their word forms. Along with the roots, the grammatical information pertaining to that particular paradigm or pattern is stored. This scheme of the morphological analyser is given below:

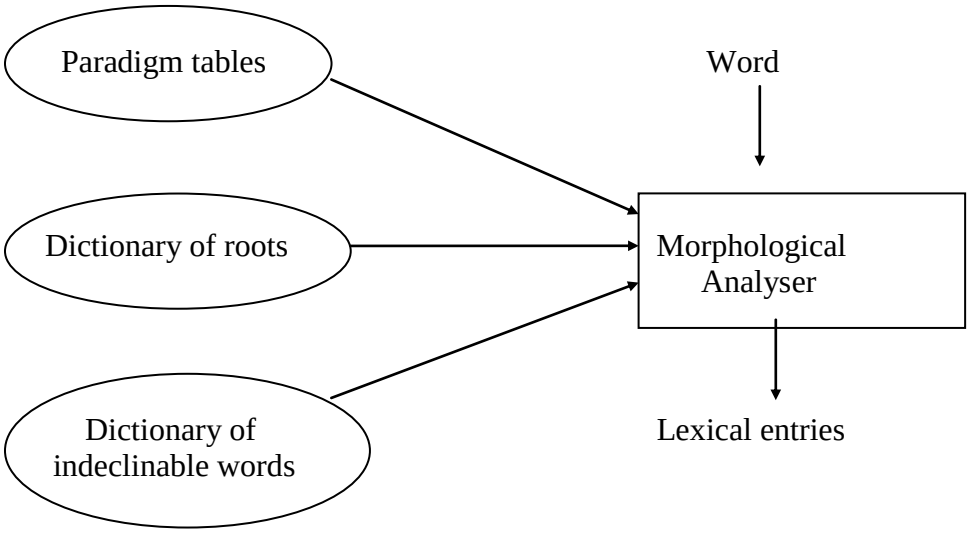


Figure: Morphological Analyser

For the purpose of spell checking we need to prepare at least the PNG (person-number-gender) analyser for verb forms and noun forms.

Analysis of Nominal complex in Tamil

Nouns undergo the morphological process of inflection, derivation and compounding. Nouns inflect mainly for number and cases. The case is followed by the postpositions which are mostly non-inflectional in their behaviour. The suffixation and the addition of postposition form a complex which we call here as nominal complex. As the postpositions are functionally too closer to case suffixes, are controlled by the case suffixes, the morphological analyzer-cum-generator discussed here take into to account the formation of case forms as well as the

postpositioal forms. The preparation of morphological analyzer for Tamil is explained below.

Case Suffixes in Tamil

All the case suffixes in Tamil starts with vowel phoneme (Example: *ai, aal, ooTu, uTan, ukku, in, uTaiya, atu, il*). Though the third case suffix is taken as *-ku* by the traditional Tamil grammarians it is better take it as *-ukku* to account for the sandhi changes. Tamil Grammarians name cases as first case, second case, and so up and they posit eight such cases. These cases are named as subject case, object case, instrumental case, sociative case, benefactive case, ablative case and locative case based on the meaning conveyed by them (Paramasivam, 1983). A case suffix can accommodate a number of case relations and more than one case suffix can show the same case relation. As we are going to analyze the word forms which are nouns, we will call the cases as *ai*-case, *aal*-case, *ooTu*-case, *ukku*-case and so on.

Sandhi Changes to Account for the Case Inflections

The nouns inflect for case suffixes the sandhi changes may occur. Such changes are listed below:

1. The nouns ending in *aa, oo* and *uu* will take a onglide *-v-* when they are inflected for case suffixes.

aa/oo/uu + Case suffixes → aa/oo/uu v Case Suffixes

Example:

kaTaa + vai > kaTaavai
kaTaa + aal > kaTaavaal
kaTaa + ooTu > kaTaavooTu
kaTaa + uTan > aTaavuTan
kaTaa + ukku > kaTavukku
kaTaa + in > kaTaavin
kaTaa + il > kaTaavil

2. The nouns ending in *i, ii, ai* will take the onglide *-y-* when inflected for case suffixes except *ukku*.

i/ii/ai + Case Suffixes except ukku → u/ii/ai y Case Suffixes

Example:

eli + ai > eliyai
eli + aal > eliyaal
eli + ooTu > eliyooTu

eli + uTan > eliyuTan
eli + in > eliyin
eli + il > eliyil

3. When the nouns ending in *i, ii, ai, y* take second case suffix *ukku, u* of the case suffix will be lost.

i/ii/ai/y + ukku > i/ii/ai/y kku

Examples:

eli + ukku > elikku
ii + ukku > iikku
yaanai + ukku > yaanaikku
ndey + ukku > ndeykku

4. Nouns of the syllabic structure CVCu (Example: pacu, maTu, maRu, teru) when inflect for case suffix will take the onglide -*v*-.

CVCu + Case Suffixes > CVCu v Case Suffixes

Examples:

pacu + ai > pacuvai
pacu + aal > pacuvaal
pacu + ooTu > pacuvooTu
pacu + uTan > pacuvuTan
pacu + ukku > pacuvukku
pacu + in > pacuvin
pacu + il > pacuvil

5. Nouns except the syllabic structue CVCu but ending with the syllablic structure (V)VTu, *u* of the noun will be dropped resulting in the doubling of *T*.

Examples:

kaaTu + ai > kaaTTai
kaaTu + aal > kaaTTaal
kaaTu + ooTu > kaaTTooTu
kaaTu + uTan > kaaTTuTan
kaaTu + ukku > kaaTTukku
kaaTu + il > kaaTTil

There are exceptions to this rule. For Example, *takaTu, utaTu* will behave like *kaaTu*, whereas *kuRaTu, kacaTu*, etc. will not behave like *kaaTu*

Examples:

kuRaTu + ai > kuRaTai
kacaTu + ai > kacaTai

6. Nouns ending in u with the syllabic structure (V)VRu when inflect for case suffixes, the final *u* of the nouns will be dropped resulting in the doubling of *R* of the last syllable.

Examples:

vayiRu + ai > vayiRRai
vuyiRu + aal > vayiRRaal
vayiRu + ooTu > vaiRRooTu
vayiRu + uTah > vaiRRuTan
vayiRu + ukku > vayiRRukku

7. Nouns ending in *u,* but with the exception of syllabic structures CVCu, (V)VTu and -(V)VRu mentioned above (Example: vaNTu, kaacu, katavu, kaTuku, etc.), the final *u* will be dropped.

Ex.

vaNTu + ai > vaNTai
vaNTu + ukku > vaNTukku
vaNTu + aal > vaNTaal
vaNTu + ooTu > vaNTooTu
vaNTu + uTan > vaNTuTan
vaNTu + il > vaNTil
vaNTu + in > vaNTin

kaacu + ai > kaacai
kaacu + ukku > kaacukku
kaacu + aal > kaacaal
kaacu + ooTu > kaacooTu
kaacu + uTan > kaacuTan
kaacu + il > kaacil
kaacu + in > kaacin

8. Nouns with the syllabic structure (C)VC ending in N, l, L or n when inflected for case suffixes(which begins with vowel), the final C will gemminate, i.e. N, l, L and n will gemminate into NN, ll, LL, and nn respectively.

N//l/L + Case Suffixes > NN/yy/ll/LL Case Suffixes

Ex.

kaN + ai > kaNNai
kaN + aal > kaNNaal
kaN+ ukku > kaNNukku
kaN+ilirundndtu > kaNNilirundtu
kaN+in > kaNNin
kaN+il >kaNNil
ndel+aal >ndellaal
kaL+ukku >kaLLukku
pon + in > ponnin

9. Nouns with the syllabic structure (C)VC ending in *y* when inflected for case suffixes except *ukku*, the final *y* will gemminate.

Ex.
ndey + ai > ndeyyai
ndey + aal > ndeyyaal
ndey + ilirundtu > ndeyyilirundtu
ndey + in > ndeyyin
ndey + il > ndeyyil

10. Nouns with the syllabic structue (C)VC ending in *y* when inflected for case suffix *ukku*, the *u* of *ukku* will be dropped.

Ex.

ndey + ukku > ndeykku

11. Nouns ending in *-am* when inflected for case suffixes, the final *m* of the nouns will be dropped with the formation of an inflectional increment *att.*

Ex.
maram + ai > marattai
maram + all > marattaal
maram + ukku > marattukku
maram + ilirundtu > marattilirundtu
maram + in > marattin
maram + il > marattil

12. Nouns ending in *a* (for example *pala*, *cila*) when inflected for case suffixes, an inflectional increment *vaRR* will be inserted in between.

a + Case Suffixes > a vaRR Case Suffixes

Ex.
pala + ai > palavaRRai
pala + aal > palavaRRaal
pala + ukku > palavaRRukku
pala + ilirindtu > palavaRRilirundtu
pala + in > palavaRRin
pala + il > palavaRRil

13. The noun *ellaam* when inflected for case suffixes, the final *m* of the noun will be dropped with the formation an inflectional increment *vaRR* in between.

ellaam + ai > ellaavaRRai
ellaam + aal. > ellavaRRaal
ellaam + ukku > ellaavaRRukku
ellaam + ilirundtu > ellaavaRRilirudntu
ellaam + in > ellaavaRRin
ellaam + il > ellavaRRil

13. The noun *cuvar* when inflects for case suffix, the final *r* of the noun will be changed into *RR*; *kaL* will be added to the noun without any change. The only member of this class is the noun *cuvar.*

cuvar + ai > cuvaRRai

14. The nouns which do not belong to the above mentioned syllabic structures, will inflect without any change.

kaal + ai > kaalai
kaal + aal > kaalaal
kaal + ukku > kaalukku
kaal + ilirundtu > kaalilirudndtu
kaal + in > kaalin
kaal + il > kaalil

Sandhi Changes due the suffixation of the plural suffix *kaL*

The following are the sandhi rules to account for the changes which takes place when nouns are inflected for the plural suffix -*kaL*.

1. When nouns ending in *aa,ii,uu* and *uu* inflect for *kaL*, the *k* of the *kaL* suffix will gemminate.

aa/ii/uu/oo/u + kaL > aa/ii/uu/oo/u kkaL

Ex.

kaTaa + kaL > kaTaakkaL
ii + kaL > iikkaL
puu + kaL > puukkaL
pooTToo + kaL > pooTTookkaL
pacu + kaL > pacukkaL

2. When suffix *kaL* is added with nouns ending in *-am*, the final *m* of the nouns will change into *ng*.

Ex.

maram + kaL > marangkaL

3. When suffix *kaL* is added with nouns of the syllabic structure (C)VC in which the final C is *l*, then *l* will change into *R*.

pul + kaL > puRkaL

4. When suffix *kaL* is added with nouns of the syllabic structure (C)VC in which the final C is *L*, then *L* will change into *T*.

Example:

muL + kaL > muTkaL

There are many exceptions to this rule. The application of this rule is rare in Modern Tamil and there are some forms in which the application of this rule is a must.

poruL + kaL > poruLkaL/poruTkaL
ndaaL + kaL > ndaaTkaL
tooL + kaL > toorLkaL/*toTkaL

5. The nouns which do not belong to the above type of syllabic structure, will inflect without any change when *kaL* is added.

Ex.
puli + kaL > pulikaL
puunai + kaL > puunaikaL
kaaTu + kaL > kaaTukaL
kaN + kaL > kaNkaL
ndaay + kaL > ndaaykaL
malar + kaL > malarkaL
itaz + kaL > itazkaL
maan + kaL > maankaL
kaal + kaL > kaalkaL

tooL + kaL > tooLkaL

The Consolidated Seventeen Classes of Nouns

Seventeen classes of nouns have to be posited to account for the sandhi changes for the sake of morphological analyzer and generator for nouns. They are listed below giving the details of the sandhi changes they undergo when inflected for case suffixes and plural number.

1. *kaTaa*-class: The nouns of *kaTaa*-class receives an onglide *v* when inflected for case suffixes. The nouns like *kaTaa, aNTaa, tukkaTaa* are included under this class. *k* of plural suffix *kaL* geminates into *kk* resulting in *kkaL*.

2. *eli*-class: The nouns of *eli*-class receives an onglide *y* when inflected for case suffixes. The plural *kaL* is added without any change. *eli, puli, yaanai* are classified here.

3. *ii*-class: The nouns of *ii*-class receives an onglide *y* when case suffixes except *ukku* are added; *ukku* will change into *kku* by dropping *u*. *k* of *kaL* will not change.

4. *vaNTu*-class: The case suffixes beginning with *u* will lose their *u* when added with nouns of *vaNTu* class. *kaL* is added without any change. The nouns such as *vaNTu, toNTu, kaacu, aaRu* (numeral noun) are classified under *vaNTu*-class of nouns.

5. *kaaTu*-class: The nouns of *kaaTu*-class will lose *u* and the penultimate consonant *T* will gemminate when case suffixes are added. *kaL* will be added without any change.

6. *aaRu*-class: The nouns of *aaRu*-class loses the final *u* and the penultimate consonant *R* geminates into *RR* when case suffixes are added. The nouns like *aaRu, vayiRu* 'stomach', *cooRu* 'rice' are grouped under this class.

7. *kaN*-class: Nouns of *kaN*-class when inflected for case suffixes, the final *N* of the nouns geminates into *NN*. The plural *kaL* is added without any change. The nouns such *kaN, viN, maN* belongs to this class.

8. *pon*-class: The nouns of *pon*-class when inflected for case suffixes, the final *n* geminates into *nn*. The plural *kaL* is added without any change. The nouns such as *pon,* etc. belong this class of nouns.

9. *ndey*-class: The nouns of *ndey*-class when added with case suffixes except *ukku,* the final *y* of the nouns will geminate; *u* of *ukku* will be dropped; *kaL* will be added without any change. The nouns such as *ndey* 'ghee', *mey* 'truth', *poy* 'falsehood' are included under *ndey*-class.

10. *ndaay*-class: The nouns of *ndaay*-class does not change when the case suffixes except *ukku* are added to them; *ukku* changes into *kku* by dropping the initial *u*. The nouns such as *ndtaay* 'dog', *paay* 'mat', *ceey* 'child' are grouped under nouns of *ndtaay* class.

11. *pul*-class: When the cases suffixes are added to the nouns of *pul*-class, the final *l* of the nouns geminates into *ll*; when *kaL* is added, *l* changes into *R*. The nouns such as *pul* 'grass', *kaL* 'tody', *pal* 'teeth' are grouped in this class.

12. *muL*-class: When case suffixes are added to the nouns of *muL*-class, the final *L* of nouns geminates into *LL*; when *kaL* is added, *L* changes into *T*. The nouns such *muL, ndtaaL, aaL* are included under this group.

13. *maram*-class: When case suffixes are added to the nouns of *maram*-class, the final *m* is dropped and an inflectional increment *att* appears in its place; *m* changes into *ng* when *kaL* is added. The nouns such as *mara*m 'tree', *karam* 'hand', *pamparam* 'back side' are grouped under *maram*-class.

14. *pala*-class: When the nouns of *pala*-class inflect for case suffixes, an inflectional increment *vaRR* appears before the case markers. *kaL* is not added to this class of nouns. *pala* 'many' *cila* 'few ones' belongs to this class.

15. *ellaam*-class: The nouns of *ellaam* class when inflect for case suffix, an inflectional increment *vaRR* will appear with the dropping of final *m*. *kaL* will not be added to this class of nouns.

16. *cuvar*-class: The nouns of *cuvar* class when inflect for case suffix, the final *r* of the noun will be changed into *RR*; *kaL* will be added to the noun without any change. The only member of this class is the noun *cuvar.*

17. *kaal*-class: The nouns of *kaal* class will inflect for case and plural without any change.

Apart form these classes, there need to include the following pronouns, as they have oblique forms (when inflected for case suffixes) which are different form their nominative forms.

Sr.No.	Nominative form	oblique form
1	ndaan 'I'	en
2	ndaam 'we (inclusive)'	ndam
3	ndaangkaL 'we (exclusive)'	engkaL
4	ndii 'you'	un
5	ndiir 'you (honorofic)'	um

6	ndiingaL ‘you (more honorific/plural)	ungaL
7	avai ‘they (neutral)’	avaRR
8	taan ‘he, she,it’	tan
9	taangkaL	tangkaL

The oblique form of *ndaan*, *en* when inflects for case suffixes behaves like *pon*-class of verbs, except that, *en* takes the dative suffix as *akku*, rather than *ukku*. Similarly The oblique form of *ndii*, *un* when inflects for case suffixes behaves like *pon*-class of verbs, except that, *un* takes the dative suffix as *akku*, rather than *ukku*. The oblique form of *ndaam*, *ndam* does not behave like *maram*. It behaves like *pon*-class of nouns, except that *pon*-class of nouns has *n* as their final consonant, where as nam has m as the final consonant. The oblique form of *ndiir*, *um* also be behaves like *pon*-class of nouns, except that *pon*-class of nouns has *n* as their final consonant, where as *um* has *m* as the final consonant. The oblique form of *ndiingkaL*, *ungkaL* behave like kaal-class of nouns, except that *ungkaL* does not take plural *kaL*. Thus we have the following, additional class of nouns.

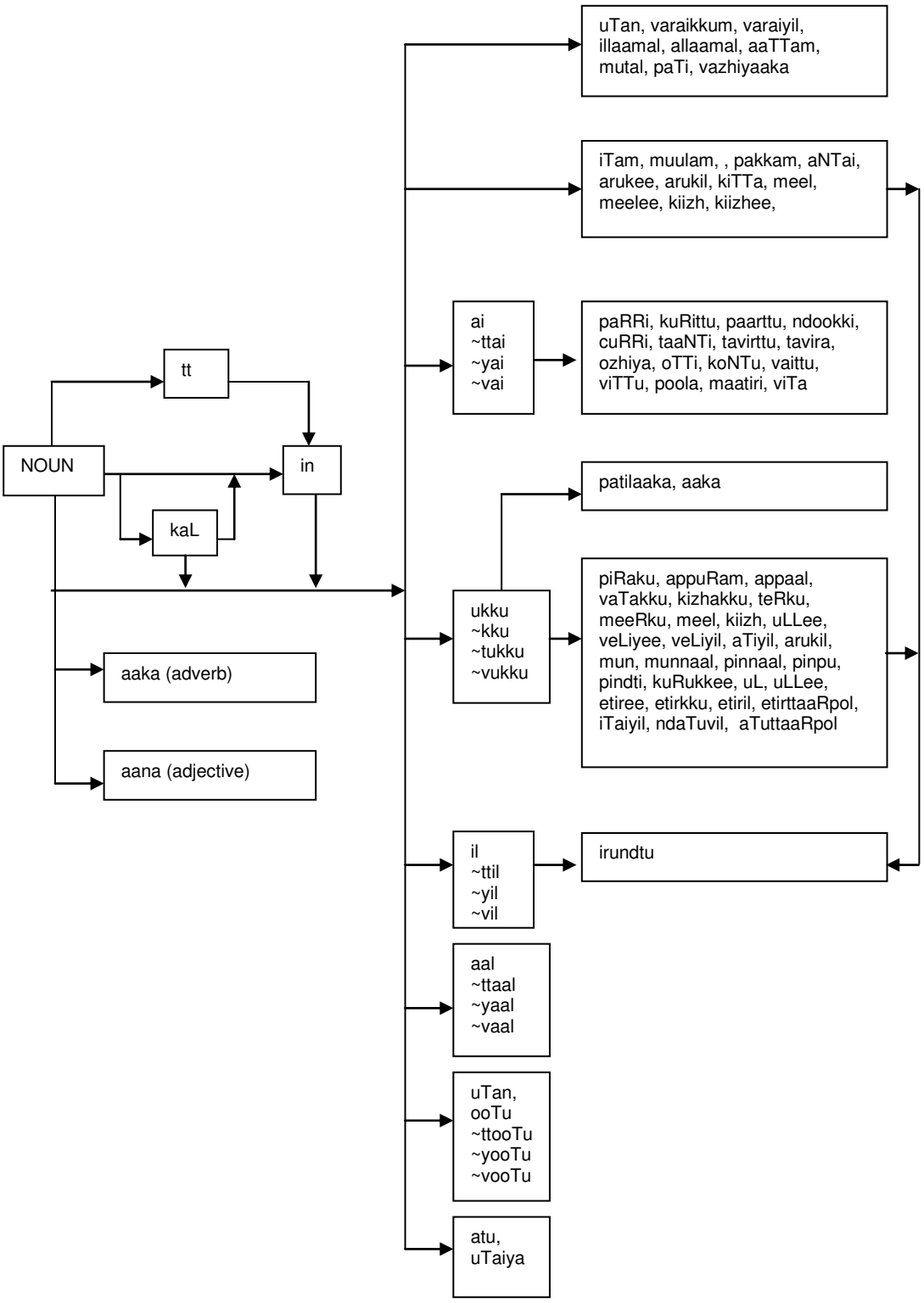
18. *ndaan*-class, 19. *ndaam*-class, 20.*ndaangkaL*-class, 21.*ndii*-class, 22. *ndiir*-class, 23.*ndiingkaL*-class, 24. *avai*-class, 25. taan-class, 26. taangkaL-class.

Structure of Nominal complex

A noun can take a plural suffix optionally followed by case suffixes. The case suffixes takes postpositions which are simple and complex in nature. The following is the structural formula of nominal complex.

Noun+Plural suffix+ Case suffix+Postposition
]
The flow chart given below shows the formation of nominal complex.

Flowchart of Tamil Noun Forms


uTan, varaikkum, varaiyil, illaamal, allaamal, aaTTam, mutal, paTi, vazhiyaaka
iTam, muulam, , pakkam, aNTai, arukee, arukil, kiTTa, meel, meelee, kiizh, kiizhee,
ai ~ttai ~yai ~vai
paRRi, kuRittu, paarttu, ndookki, cuRRi, taaNTi, tavirttu, tavira, ozhiya, oTTi, koNTu, vaittu, viTTu, poola, maatiri, viTa
tt
NOUN
in
kaL
patilaaka, aaka
ukku ~kku ~tukku ~vukku
piRaku, appuRam, appaal, vaTakku, kizhakku, teRku, meeRku, meel, kiizh, uLLee, veLiyee, veLiyil, aTiyil, arukil, mun, munnaal, pinnaal, pinpu, pindti, kuRukkee, uL, uLLee, etiree, etirkku, etiril, etirttaaRpol, iTaiyil, ndaTuvil, aTuttaaRpol
aaka (adverb)
aana (adjective)
il ~ttil ~yil ~vil
irundtu
aal ~ttaal ~yaal ~vaal
uTan, ooTu ~ttooTu ~yooTu ~vooTu
atu, uTaiya

Tabulation of the flow of suffixes

The tables given below show the flow of suffixes form one level to another level and the set of suffixes occurring in each level.

Table showing the first level base and second level of suffixation

Noun class	Base alternatns	Plural suffix	Accusative case suffix	Instrumental case suffix	Dative case suffix	Genitive case suffix-1	Genitive case suffix-2	Genitive case suffix-3	Locative case suffix	Sociative case suffix	Sociative case suffix
class-1 *kaTaa*-class	kaTaa ~ kaTaav	kkaL	ai/inai	aal/inaal	ukku/inukku	in	atu/ inatu	uTaiya/inuTaiya	il/inil	ooTu	uTan
class-*2 eli*-class	eli ~ eliy	kaL	ai/inai	aal/inaal	ukku/inukku	in	atu/ inatu	uTaiya/inuTaiya	il/inil	ooTu	uTan
class 3 ***ii*-class**	ii~iiy	kkaL	ai/inai	aal/inaal	ukku/inukku	in	atu/ inatu	uTaiya/inuTaiya uTaiya/inuTaiya	il/inil	ooTu	uTan
class 4 *vaNTu*-class	vaNTu ~ vaNT	kaL	ai/inai	aal/inaal	ukku/inukku	in	atu/ inatu	uTaiya/inuTaiya	il/inil	ooTu	uTan
class 5 *kaaTu*-class	kaaTu ~ kaaTT	kaL	ai/inai	aal/inaal	ukku/inukku	in	atu/ inatu	uTaiya/inuTaiya	il/inil	ooTu	uTan
class 6 *aaRu*-class	aaRu ~aaRR	aaRukaL	ai/inai	aal/inaal	ukku/inukku	in	atu/ inatu	uTaiya/inuTaiya	il/inil	ooTu	uTan
class 7 *kaN*-class	kaN ~ kaNN	kaL	ai/inai	aal/inaal	ukku/inukku	in	atu/ inatu	uTaiya/inuTaiya	il/inil	ooTu	uTan
class 8 *pon*-class	pon ~ ponn	kaL	ai/inai	aal/inaal	ukku/inukku	in	atu/ inatu	uTaiya/inuTaiya	il/inil	ooTu	uTan
class 9 *ndey*-class	ndey ~ ndeyy	kaL	ai/inai	aal/inaal	kku/inukku	in	atu/ inatu	uTaiya/inuTaiya	il/inil	ooTu	uTan
class 10 *ndaay*-class	ndaay	kaL	ai/inai	aal/inaal	kku/inukku	in	atu/ inatu	uTaiya/inuTaiya	il/inil	ooTu	uTan
class 11 *pul*-	pul ~ pull~puR	kaL	ai/inai	aal/inaal	ukku/inukku	in	atu/ inatu	uTaiya/inuTaiya	il/inil	ooTu	uTan

class											
class 12 *muL*-class	muL ~ muLL ~muT	kaL	ai/inai	aal/inaal	ukku/inukku	in	atu/ inatu	uTaiya/inuTaiya	il/inil	ooTu	uTan
class 13 *maram*-class	maram ~ marang~ maratt	kaL	ai/inai	aal/inaal	ukku/inukku	in	atu/ inatu	uTaiya/inuTaiya il/inil	il/inil	ooTu	uTan
class 14 *pala*-class	pala ~ palavaRR	-	ai/inai	aal/inaal	ukku/inukku	in	atu/ inatu	uTaiya/inuTaiya	il/inil	ooTu	uTan
class 15 *ellaam*-class	ellaam ~ ellaavaRR	-	ai/inai	aal/inaal	ukku/inukku	in	atu/ inatu	uTaiya/inuTaiya	il/inil	ooTu	uTan
class 16 *cuvar*-class	cuvar ~cuvaRR	kaL	ai/inai	aal/inaal	ukku/inukku	in	atu/ inatu	uTaiya/inuTaiya	il/inil	ooTu	uTan
class 17 *kaal*-class	kaal	kaL	ai/inai	aal/inaal	ukku/inukku	in	atu/ inatu	uTaiya/inuTaiya	il/inil	ooTu	uTan
class 18 *ndaan*-class	ndaan ~en~ enn	-	ai/inai	aal/inaal	ukku/inukku	in	atu/ inatu	uTaiya/inuTaiya	il/inil	ooTu	uTan
class 19 *ndaam*-class	ndaam ~ndam ~ndamm	-	ai/inai	aal/inaal	ukku/inukku	in	atu/ inatu	uTaiya/inuTaiya	il/inil	ooTu	uTan
class 20 *ndaangkaL*-class	ndaangkaL~ engkaL	-	ai/inai	aal/inaal	ukku/inukku	in	atu/ inatu	uTaiya/inuTaiya	il/inil	ooTu	uTan
class 21 *ndii*-class	ndii~ un~ unn	-	ai/inai	aal/inaal	ukku/inukku	in	atu/ inatu	uTaiya/inuTaiya	il/inil	ooTu	uTan
class 22 *ndiir*-class	ndiir~ um~ umm	-	ai/inai	aal/inaal	ukku/inukku	in	atu/ inatu	uTaiya/inuTaiya	il/inil	ooTu	uTan
class 23 *ndiingaL*-class	ndiingkaL~ ungkaL	-	ai/inai	aal/inaal	ukku/inukku	in	atu/ inatu	uTaiya/inuTaiya	il/inil	ooTu	uTan
class 24 *taan*-	taan~ tan~ tann	-	ai/inai	aal/inaal	ukku/inukku	in	atu/ inatu	uTaiya/inuTaiya	il/inil	ooTu	uTan

class											
class 25 *taangkaL*-class	taangkaL~tangkaL	-	ai/inai	aal/inaal	ukku/inukku	in	atu/inatu	uTaiya/inuTaiya	il/inil	ooTu	uTan
class 26 *avai*-class	avai~avaRR	-	ai/inai	aal/inaal	ukku/inukku	in	atu/inatu	uTaiya/inuTaiya	il/inil	ooTu	uTan

Table showing the second/third level of

Set of postpositions occurring immediately after the nominative forms	**Set of postpositions occurring immediately after the oblique forms or after the *in* suffixed oblique forms**	**Set of postpositions occurring after the accusative case suffix *ai***	**Set of postpositions occurring after the dative case suffix *ukku***	**The postposition occurring after the locative case suffix *il* and other star (*) marked postpositions**	
uTan	iTam*	paTTi	pakkam*	irundtu	
varai	pakkam*	kuRittu	pakkattil*		
varaikkum	pakkattil*	oTTi	aNTai*		
illaamal	aNTai	paarttu	aNTaiyil*		
muulam	aNTaiyil*	ndookki	arukee*		
aaka	arukee*	cuRRi	arukil*		
aana	arukil*	taaNTi	aTuttaaRpoola*		
	aTuttaaRpoola	tavirttu	kiTTee*		
	kiTTee*	tavira	meel*		
	meel*	oziya	meelee*		
	meelee*	koNTu	kiiz*		
	kiiz*	vaittu	kiizee*		
	kiizee*	viTTu	uL*		
	uL*	poola	uLLee*		
	uLLee	maatiri	veLiyee*		
	veLiyee*	viTa	veLiyil*		
	veLiyil*		mun*		
	mun*		munnaal*		
	munnaal*		pinnaal		
	pinnaal		pinpu*		
	pinpu*		pindti		
	pindti		kuRukkee*		
	etiree*		etiree*		
	etiril*		etiril*		
	etirttaaRpool*		etirttaaRpool*		
	iTaiyil*		iTaiyil*		
	ndaTuvil*		ndaTuvil*		
	kuRukkee*		piRaku*		
			appuRam*		
			appaal*		
			vaTakku		

			vaTakkee*		
			vaTakkil*		
			kizakku		
			kizakkee*		
			kizakkil*		
			teRku		
			teRkee*		
			teRkuil*		
			meeRku		
			meeRkee*		
			aaka		
			veeNTi		
			patil		
			patilaaka		

Analysis of verbal complex

The following are the major issues to be faced:

- What are the morphophonemic or sandhi rules which account for the shape changes in the formation of verbal complex?
- How the verbs are combined and how the meanings of the verbal complex are interpreted based on the combination of meanings of the constituents of the verbal complex?
- What are the possible and impossible permutations?
- How to restrict the generation of ungrammatical verbal complex and unlawful meaning interpretation?

Morphological analysis (parsing or recognition) is concerned with retrieving the structure, the syntactic and morphological properties, or the meaning of a morphologically complex word. Morphological generation is concerned with producing an appropriate morphologically complex word from some set of semantic or syntactic features. We need to understand the following things:

1. What kinds of representations a morphological processing system must deal in before we can appreciate the issues involved in designing such a system, or give a fair evaluation of the performance of a particular morphological processing system?
2. What sort of things morphology exhibit in a language?
3. How are words built up form smaller meaningful units – morphemes?
4. What are the constraints on the order of morphemes within words?
5. Do phonological rules complicate the problem of morphological analysis?

Theoretical morphologists portray word structure as being very intricate and complex, whereas computational morphologists have built working system that deal with large fragments of the morphology of languages in which the morphology is

seemingly quite complex. No existing processing system can handle the entire range of possibilities or morphological expression in natural language. While a theoretical morphologist is interested in assigning phonological elements to morphs and realizing surface realties into abstract properties, a computational morphologists, on the other hand may be interested in more immediate goals, such as providing a working system which will handle 99 percent of the words encountered in texts.

The present paper aims to address the strategies to be adopted and the linguistic and computational issues to be faced while building a tool which aims to analyze and generate the verbal complex of Tamil. Broadly speaking, the task at our hand can be divided into two:

1. Understanding and tackling of verbal complexity
2. Computational analysis and generation of the verbal complexes

Understanding and tackling of verbal complexity

The understanding of verbal complexity involves understanding the structure of simple finite verbs and compound verbs from the point of view of analysis and generation and the characteristic features of compound verbs. By understanding the nature of the verbal complexity, it is possible to evolve a methodology to tackle the verbal complexity.

Composition of a finite verb

A simple Tamil finite verbal form consists of a verbal root followed by a tense marker or negative marker and person-number-gender markers.

Verb + Tense/Negative+PNG
cey-t-aan ‘do-PAST-he’
cey-aa-tu ‘do-NEG-it’

Composition of compound verbs

A compound verb consists of a non-finite verb followed by one or more auxiliary verbs. The verbal pattern can be generalized as follows:

Verb+(Tense/Negative)+Nonfinite suffix + Auxiliary Verb....Auxiliary Verb+ Tense/Negative + PNG

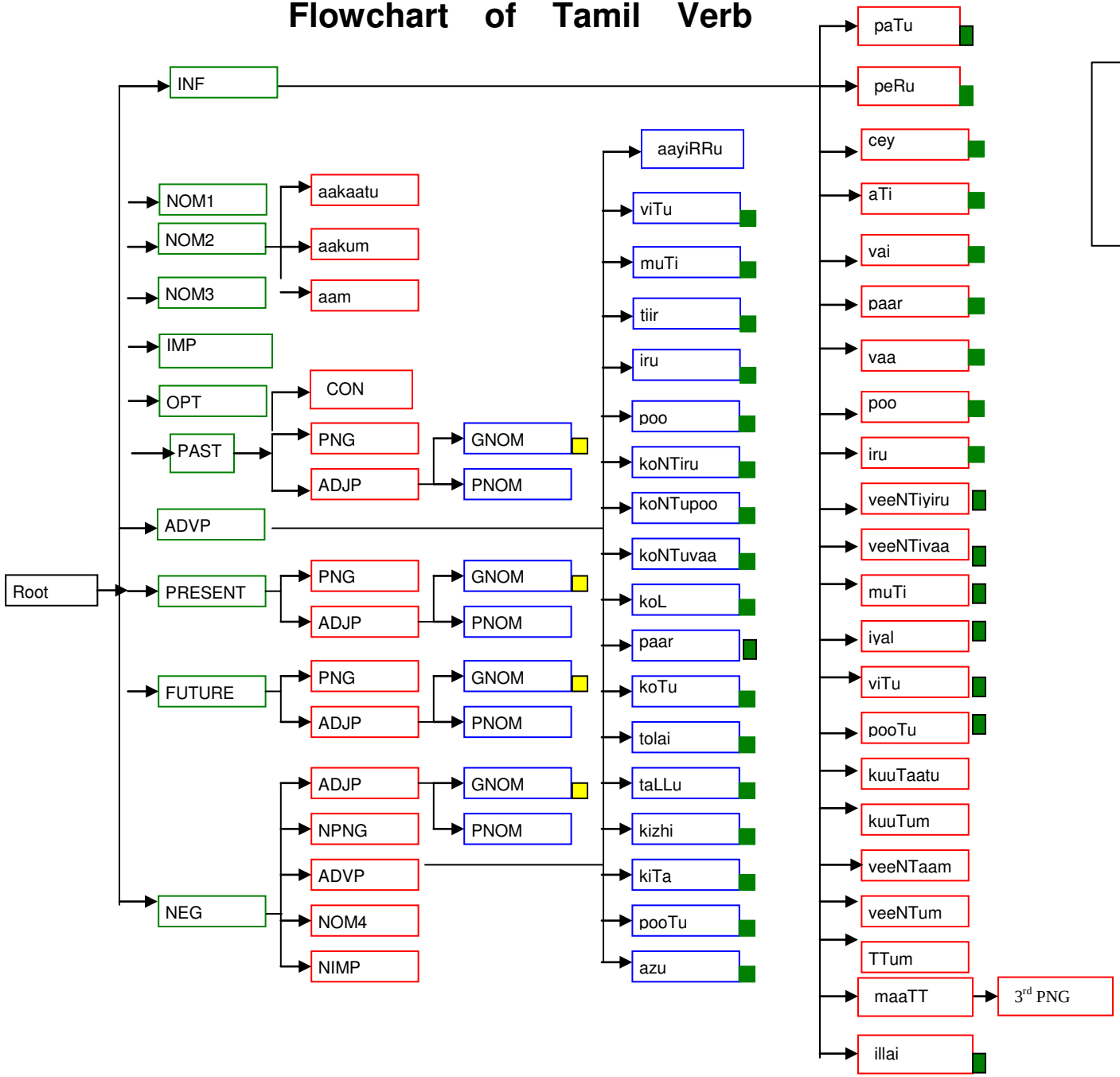
Ex.
cey-t-u viT-T-aan ‘do-PAST-ADVP leave-PAST-he’
‘he did it (completive sense is emphasized)’

The non-finite forms (which include certain type of verbal noun forms) of a verb can be of at least five types:

Type	**Structural description**	**Example**
1. Adverbial form	Verb + Adverbial marker	*cey-tu* ‘having done'
“	Verb + Negative Adverbial marker	*cey-aatu* ‘having not done'
2. Adjectival form	Verb + Tense/Negative + Adjectival marker	*cey-t-a* ‘that which is done' *ceyy-aat-a* ‘that which is not done’
3. Infinitive form	Verb + Infinitive marker	*ceyy-a* ‘to do’
4.1. *al*-Nominlized verbal form	Verb+ Nominalizer *al*	*ceyy-al* ‘doing'
4.1. *atu*-Nominlized verbal form	Verb+Tense/Negative+Nominalizer *atu*	*cey-t-atu* ‘that which is done'
5. Conditional form	Verb + PAST-*aal*	*cey-t-aal* 'if done'
“	Verb + Negative + *viT*-PAST-*aal*	*ceyy-aa-viT-T-aal* ‘if not done'

The auxiliary verbs are added to the adverbial forms, infinitive forms, *al*-nominalized forms and *atu*-nominalized forms. The following flow chart depicts the formation of all the verbal forms in Tamil.

Flowchart of Tamil Verb



PRESENT : [kiR ~kkiR], [kinR ~kkinR]
PAST : t, ndt, NT, nR, in ~n ~nn, R, T, tt
FUTURE : p, pp, v, um

IMP : um, umeen, ungkaL, ungkaLeen, een
INF : a ~ka ~kka, OPT: ka
NEG : aa~aat, NIMP : ee, CON: aal

NOM1: al ~kal ~kkal
NOM2: tal ~ttal ~Ral
NOM3: kai~kkai

NOM4: mai, ADVP : [tu ~ntu ~NTu~nRu~Tu~ Ru~ ttu~i ~y], [mal~tu],
ADJP : a~ ∅], PNG : een, oom, aay, iir, iirkaL, aan, aaL, aar, aarkaL, atu, ana
PNOM : van, vaL, var, varkaL, tu, vai, na , GNOM: tu, vai, NPNG: tu

Characteristic features of the compound verbs

The compounds show certain characteristic features that distinguish them from phrases or conjoined verbs.

1. The constituents of a verbal compound cannot be intervened by other words, whereas the conjoined verbs allow other words to intervene between them.

2. The order of the constituents in a compound is fixed. The cohesive nature of the compound will be lost if we resort to any reversal of the order or displacement of the constituents of the compound.

3. The constituents of a compound are combined into a single unit and so they are referentially opaque. It is not possible to factor out the individual constituents of a compound and apply a syntactic operation on it as the compound verbs function as simple words showing lexical integrity.

4. The compound verbs are not derived through clausal embedding, whereas conjoined verbs are derived through clausal embedding.

5. Hook (1974:94) uses grammaticalization as one of the discriminating properties of compound verbs. The same phenomenon can be adopted for Tamil and other Dravidian languages too. The V2 of compounds lose their primary lexical meaning, whereas the V2 of conjoined verbs retain their primary lexical meaning.

Ex.
avan poo-n-aan
he go_past_he
'He went'
avan ce-ttup poo-n-aan
'he die_advp go_past_he
'He passed away'

6. When negative verb *illai* is added to V1-V2 complex, it negates the entire verbal complex (i.e. the main verb), whereas in the case of conjoined verbs it negates only the verb to which it is added.

Ex.
avan ce-ttup poo-ka-villai
he die_advp go_inf_not
'He did not pass away'

avan uNavu caapiT-Tu viT-Tup paTi-kka-villai
he food eat_advp leave_advp study_not
'After having eating the meals, he did not study'

Classification of the auxiliary verbs based on their function

The following chart shows the classification of auxiliary verbs based on their function.

Type	**Semantic features**	**Constraint on the main verb**	**Verbal form**
Argument structure changing verbs	Cause		V+ infinitive suffix + *cey*
“	Cause		V+ infinitive suffix + *vai*
“	Cause	with a set of intranstive verbs such as *caa* ‘die’, *tiNaRu* ‘suffocate’	V+ infinitive suffix + *aTi*
“	Allow	with intransitive verbs	V+ infinitive suffix + *viTu*
“	Allow	with intransitive verbs denoting change of state such as *kaay* ‘dry’, *ndanai* ‘become wet’	V+ infinitive suffix + *pooTu*
Voice changing verbs	Passive	with transitive verbs which entails change of state of the object or impact on the object	V + infinitive suffix + *paTu*
“	Passive	with transitive verbs which entails change of state of the object or impact on the object	V + infinitive suffix + *peRu*
“	Reflexive	with transitive verbs denoting impact	V + adverbial suffix + *koL*
“	Reciprocal	with transitive verb with plural subject	V + adverbial suffix + *koL*
“	Repetitive		V+ adverbial suffix + *vaa*
Aspectual verbs	Completion Perfect		V+ adverbial suffix + *aayiRRu*
“	Perfect		V+ adverbial suffix + *viTu*
“	Perfect (The impact of the action is still felt.)		V+ adverbial suffix + *iru*
“	Completion Perfect	with intransitive verbs denoting	V+ adverbial suffix + *poo*

	(Intensifies the completion and swiftness of the action.)	change of state	
“	Completion Perfect		V+ adverbial suffix + *tiir*
“	Completion Perfect (Lends force or intensity to the action denoted by the main verb.)	with transitive verbs entailing change of state of the object such as *uTai* ‘break’, *kizi* ‘tear’, etc.	V+ adverbial suffix + *eRi*
“	Imperfect Progressive	with verbs such as *uTkaar* ‘sit’ , *paTu* ‘lie down’	V+ adverbial suffix + *iru*
“	Imperfect Progressive		V+ adverbial suffix + *koNTiru*
“	Imperfect Progressive		V+ adverbial suffix + *koNTupoo*
“	Imperfect Progressive		V+ adverbial suffix + *koNTuvaa*
“	Imperfect Progressive (Denotes that the state described by the main verb remains undisturbed.)		V+ adverbial suffix + *kiTa*
Model auxiliaries	Inceptive		V+ infinitive suffix + *poo*
“	Attemptive (Denotes that the action expressed by the main verb is done as an attempt or trail or to see what it leads to.)		V + adverbial suffix + *paar*
“	Abilitative (Denotes the capability or the potentiality on the part of the subject to perform the action denoted by the main verb.)		V + infinitive + *muTiyum*
“	Abilitative (Denotes the capability or the		V + infinitive + *iyalum*

	potentiality on the part of the subject to perform the action denoted by the main verb.)		
“	Inability		V + infinitive + *muTiyaatu*
“	Inability		V + infinitive + *iyalaatu*
“	Possibility		V+ infinitive + *kuuTum*
“	Probabilitative		V-*al* + *aam*
“	Permissive		V-al + aam
“	Prohibitive		V+ infinitive + *kuuTaatu*
“	Prohibitive		V-*al* + *aakaatu*
“	Obligative (Denotes the obligatoriness of the action denoted by the main verb)		V+ infinitive + *veeNTum*
	Dispensative (Denotes the non necessity of the action denoted by the main verb.)		V+ infinitive + *veeNTiyatillai*
“	Denial		V+ infinitive + *maaTTu*
“	Accelerative (Denotes the accelerated phase or rate in which the action denoted by the main verb is performed.)		V+ past participle+ *taLLu*
“	Denotes that the action is carried out or performed as a fore thought.		V+ past participle + *vai*
“	Benefactive (Denotes that the action is carried out for the benefit of somebody.)		V + past participle + *koTu*
“	Contemptive-1 (Denotes the negative attitude on the speaker towards the action		V + past partciple + *tolai*

	expressed by the main verb.)		
“	Contemptive-2 (Denotes the negative opinion of the speaker regarding the accomplishment of the action denoted by the main verb.)		V + past partciple + *kizi*
“	Contemtive-3		V + past partciple + *azu*

2.5. Classification of auxiliary verbs based on the non-finite form to which they are added

The following chart chows the classification of auxiliary verbs based on the non-finite form to which they are added.

Type of Non-finite form to which auxiliary verbs are added	**The auxiliary verbs added after the Non-finite forms**
The auxiliary verbs added after adverbial form	aayiRRu, iru, kaaTTu, kiTa, kizi, koTu, koNTiru, koL, tiir, taLLu, tolai, paar, poo, pooTu, muTi, viTu
The auxiliary verbs added after infinitive form	aTTu, iru, iyalum, illai, kuuTum, kuuTaatu, cey, paar, poo, maaTTu, muTi, muTiyum, vaa, veeNTaam veeNTiyiru, veeNTi vaa, veeNTum
The auxiliary verbs added after *al*-nominalized form	aam, aaku, aakaatu
The auxiliary verb added after *atu*-nominalized firm	uNTu, illai, kiTaiyaatu

Multiple Natures of Verbal forms

Accounting the multiple natures of verbal forms economically is the major problem in the computational analysis and generation of verbal complex in Tamil. The simplest non-economical way is to list down all the possible combinatory forms (which runs into thousands if we take into account the addition of auxiliary verbs. But the non-economical way is against the interest of a linguist. A simple Tamil verb has a primary imperative form (which is normally the citation form for verbs), other imperative/permissive forms, finite and non-finite forms, pronominalized forms, regular verbal noun forms, etc. The following table shows the list of the paradigms of verb *cey* `to do’ keeping aside the combinatory forms with auxiliary verbs.

Type of forms	**Forms of *cey***
Imperative/ Permissive forms	1. cey ‘do’ 2. ceyy-ungkaL ‘do_you (honorific)’ = ‘you (honorific) do’ 3. ceyv-iir ‘do_you’ = ‘You do’ 4. ceyy-al-aam ‘do_NOM INATIVE_may =‘You may do’ 5. cey-ka ‘do_HORTATIVE’ = ‘Please do’ 6. ceyy-aTTum ‘do_let’ = ‘Let him/she/it do’ 7. ceyy-aatee ‘do_not’ = ‘Don’t do’
Finite forms	1. cey-t-een ‘do_PAST_I’ = ‘I did’ 2. cey-kiR-een ‘do_PRESENT_I’ = ‘I do’ 3. cey-v- een ‘do_FUTURE_I = ‘ I will do’ 4. cey-t-oom ‘do_PAST_we = ‘We did’ 5. cey-kiR-oom ‘do_PRESENT_we = ‘We do’ 6. cey-v-oom ‘do_FUTURE_we = ‘We will do’ 7. cey-t-aay ‘do_PAST_you = ‘You did’ 8. cey-kiR-aay ‘do_PRESENT_you = ‘You do’ 9. cey-v-aay ‘do_FUTURE_you = ‘You will do’ 10.cey-t-iir ‘do_PRES_you = ‘You (hon.) did’ 11. cey-kiR-iir ‘do_PRES_you = ‘You (hon.) do’ 12. cey-v-iir ‘do_FUTU_you = ‘You (hon.) do’ 13. cey-t-iirkaL ‘do_PAST_you = ‘You (hon.) do’ 14. cey-kiR-iirkaL ‘do_PAST_you = ‘You (hon.) did’ 15. cey-v-iirkaL ‘You (honorific) will do’ 16. cey-t-aan ‘do_PAST_he’ = ‘He did’ 17. cey-kiR-aan ‘do_PRESENT_he = ‘He does’ 18. cey-v-aan ‘do_FUTURE-he = ‘He will do’ 19. cey-t-aaL ‘do_PAST_she = ‘She did’ 20. cey-kiR-aaL ‘do_PRESENT_she = ‘She does’ 21. cey-v-aaL ‘do_FUTU_she = ‘She will do’ 22. cey-t-aar ‘do_PAST_he/she = ‘He/She did’ 23. cey-kiR-aar ‘do_PRESENT_he/she = ‘He/she does’ 24. cey-v-aar ‘do_FUTURE_he/she = ‘He/she will do’ 25. cey-t-aarkaL ‘do PAST_they = ‘They did’ 26. cey-kiR-aarkaL ‘do_PRESENT_they = ‘They do’ 27. cey-v-aarkaL ‘do_FUTURE_they = ‘They will do’ 28. cey-t-atu ‘do_PAST_it = ‘It will do’ 29. cey-kiR-atu ‘do_PRESENT_It = ‘it does’ 30. ceyy-um ‘do_FUTURE_it = ‘It will do’ 31. cey-t-ana ‘do_PAST_they = ‘They (neuter) did’ 32. cey-kinR-ana ‘do_PRESENT_they (n) = ‘They (neuter) do’ 33. ceyy-um ‘no_they (neuter)’ = ‘It/They (neuter) will do’
Nonfinite forms	1. cey-tu ‘do_ADVERBIAL PARTICIPLE’ 2. ceyy-aamal ‘do_NEGATIVE ADVERBIAL PARTICIPLE’ 3. ceyy-aatu ‘do_NGATIVE ADVERBIAL PARTICIPLE’ 4. ceyy-a ‘do_INFINITIVE’ 5. cey-t-a ‘do_PAST_ADJECTIVAL PARTICIPLE’ 6. cey-kiR-a ‘do_PRESENT_ ADJECTIVAL PARTICIPLE’ 7. ceyy-um ‘do_FUTURE_ ADJECTIVAL PARTICIPLE’ 8. cey-aat-a ‘do_NEGATIVE_ ADJECTIVAL PARTICIPLE’ 9. cey-t-atu ‘do_ PAST_NOMINATIVE

	10. cey-kiR-atu ‘do_PRESENT_NOMINATIVE 11. cey-v-atu ‘do_FUTURE_NOMINATIVE 12. cey-aat-atu ‘do_NEGATIVE_NOMINATIVE 11. cey-t-aal ‘do_NOMINATIVE
Conditional forms	1.cey-t-aal ‘do_PAST_CONDITIONAL’
Pronominalized forms	1. cey-t-a-van ‘do_PAST_he 2. cey-kiR-a-van ‘he_PRESENT_he 3. cey-p-a-van ‘do_FUTURE_he 4. cey-aat-a-van ‘do_PAST_he 5. cey-t-a-vaL ‘do_PAST_ ADJECTIVAL _she 6. cey-kiR-a-vaL ‘do_PRESENT_ADJECTIVAL_she’ 7. cey-p-a-vaL ‘do_FUTURE_ ADJECTIVAL_she’ 8.cey-aat-a-vaL‘do_NEGATIVE_ ADJECTIVAL _she’ 9. cey-t-a-var ‘do_PAST_he/she (honorific)’ 10. cey-kiR-a-var ‘do_PRESENT_he/she’ 11. cey-p-a-var ‘do_FUTURE_ADJECTIVAL_he/she(honorific)’ 12. cey-aat-a-var ‘do_NEGATIVE_ADJECTIVAL_he/she (honorific)’ 13. cey-t-a-varkaL ‘do_PAST_ ADJECTIVAL_they’ 14. cey-kiR-a-varkaL ‘do_PRESENT_ ADJECTIVAL_they’ 15. cey-p-a-varkaL ‘do_FUTURE ADJECTIVAL_they’ 16. ceyy-aat-a-varkaL ‘do_NEGATIVE_ ADJECTIVAL_they’ 17. cey-t-a-tu ‘do_PAST_ ADJECTIVAL_it’ 18. cey-kiR-a-tu ‘do_PRESENT_ ADJECTIVAL_it’ 19. cey-v-a-tu ‘do_FUTURE_ ADJECTIVAL_it’ 20. ceyy-aat-a-tu ‘do_NEGATIVE_ADJECTIVAL_it’ 21. cey-t-a-na ‘do_PAST_ ADJECTIVAL_they’ 22. cey-kinR-ana ‘do_PRESEENT_ ADJECTIVAL_they’ 23. cey-v-a-na ‘do_FUTURE_ ADJECTIVAL_they’ 24. ceyy-aat-a-na ‘do_NEGATIVE_ ADJECTIVAL_they’
Regular verbal noun forms	1. ceyy-al ‘do_NOMINATIVE’ 2. cey-tal ‘do_NOMINATIVE’ 3. cey-kai ‘do_NOMINATIVE’ 4. ceyy-aa-mai ‘do_NOMINATIVE’

If we calculate the possible forms of each verb it will be as follows:

Imperative/Permissive forms	- 7
Finite forms	- 33
Non-finite forms	- 11
Conditional form	- 1
Pronominalized forms	- 24
Regular Verbal noun forms	- 4

Total	80

Thus we will have at least 80 paradigms to be accounted for each verb. This number will multiply if we take into account the compound verb formation by adding auxiliary verbs with the main verbs. Storing all these forms for the 3012 verbs listed

in KTTA (*Kiriyaavin taRkaalati tamiz̲ akaraati*) in the memory of computer is a waste. And more over we need to generate these forms from verb stems and suffixes. As the tense inflection is not regular and is not phonologically conditioned, we need to classify the verbs into certain classes to capture the inflectional process.

Tackling of the verbal complexity

In order to tackle the analysis and generation of the verbal forms in which the inflection vary from one set of verbs to another, a classification of Tamil verbs based on tense inflections need to be evolved. The inflection includes finite, infinite, adjectival, adverbial and conditional forms of verbs. For the sake of computation we have to make twentyeight classes of verbs and group the verbs under each group. For the sake of convenience of remembering them along with the morphophonemic changes involved in the inflectional and derivational processes, we can label them as *cey*-class, *azu*-class, *caa*-class, *viTu* class, *peRu*-class, *pukaz*-class, *vizu*-class, *aRi*-class, *ndoo*-class, *vee*-class, *vaa*-class, *akal*-class, *kol*-class, *aaL*-class, *koL*-class, *ooTu*-class, *col*-class, *poo*-class, *kal*-class, *keeL*-class, *puuN*-class, *uN*-calss, *kaaN*-class, *tin*-class, *en*-class, *ndil*-class, *paTi*-class, and *ndaTa*-class. The following morphophonemic information on these classes of verbs will show the need for classifying the verbs into 28 classes. (The preference is given here for the item arrangement model rather that item and process model.)

1. *cey*-class: The verbs of the *cey* class take *-t-* as past tense marker, *kiR ~ kinR* as present tense marker, *v ~ um* as future tense maker and *tu* as the adverbial participle marker. The final consonant of the verbs geminates when the suffixes beginning with a vowel such as *um, a, aat, al* are added to them.

2. *az̲u*-class: The verbs of *az̲u*-class take *-t-* as past tense marker, *kiR ~ kinR* as present tense marker, *v ~ um* as future tense marker and *tu* as the adverbial participle marker. The final *u* of the verbs is deleted when the suffixes beginning with vowel such as *um, a, aat, al* are added to them.

3. *caa*-class: The verb of *caa*-class takes *t* as past tense marker, *kiR ~ kinR* as present tense marker, *v ~ um* as future tense marker and *ttu* as the adverbial participle marker. The irregular verb *caa* is the only verb belonging to this class. *caa* changes into *ce* before past tense marker *t* and adverbial marker *ttu*. An increment *k* is added in between when the suffixes beginning with vowel such as *um, a, aata, al* are added to the verb stem directly.

4. *viTu*-class: The verbs of *viTu*-class take *T* (i.e. *Tu*+t>*TT)* as the past tense marker, *kiR ~ kinR* as present tense marker, *v ~ um* as future tense marker and *Tu* (i.e. *Tu+tu>TTu)* as the adverbial participle marker. The final vowel *u* of the verbs is dropped when past tense suffix and the adverbial suffix as well as the suffixes beginning with vowel such as *um, a, aat, al* are added to them.

5. *peRu*-class: The verbs of *peRu* class take *R* (i.e. *Ru+t>RR)* as the past tense marker, *kiR ~ kinR* as present tense marker, *v ~ um* as the future tense marker and *Ru* (i.e. *Ru+tu>RRu)* as the adverbial participle marker. The final vowel *u* is dropped when past tense suffix and the adverbial suffix as well as the suffixes beginning with vowel such as *um, a, aat, al* are added to the verb stem.

6. *pukaz*-class: The verbs of *pukaz*-class take *ndt* as the past tense marker and *kiR~kinR* as present tense marker, *v ~ um* as the future tense marker and *ndtu* as the adverbial participle marker. No change in the phonemic make up of the stems is effected when the verbs of this type inflect.

7. *vizu*-class: The verbs of *vizu* class take *ndt* as the past tense marker, *kiR ~ kinR* as present tense marker, *v ~ um* as the future tense marker and *ndtu* as the adverbial participle marker. The final *u* of the verb is dropped, when the suffixes beginning with vowel such as *um, a, aat, al* are added to the verb directly.

8. *aRi*-class: The verbs of *aRi*-class take *ndt* as the past tense marker, *kiR~kinR* as present tense marker, *v ~ um* as the future tense marker and *ndtu* as the adverbial participle marker. An onglide *y* is inserted when the suffixes beginning with vowel such as *um, a, aat, al* are added directly to the verbs.

9. *ndoo*-class: The verbs of *ndoo* class take *ndt* as the past tense marker and *kiR~kinR* as present tense marker, *v ~ um* as the future tense marker, and *ndtu* as the adverbial participle marker. The verb stem *ndoo* changes into *ndo* when past tense marker and adverbial participle marker are added to the stem. When the suffixes beginning with vowel such as *um, a, aat, al* are added to the verb stem directly, an increment *k* is added in between.

10. *vee*-class: The verbs of *vee*-class take *ndt* as the past tense marker and *kiR~kinR* as present tense marker, *v ~ um* as the future tense marker, and *ndtu* as the adverbial participle marker. The verb stem *vee* changes into *ve* when past tense marker and adverbial participle marker are added to the stem. When the suffixes beginning with vowel such as *um, a, aat, al* are added to the verb stem directly, an increment *k* is added in between.

11. *vaa*-class: (*varu* is taken as the base form rather than the citation form *vaa*). The verbs of *vaa*-class take *ndt* as the past tense marker, *kiR~kinR* as present tense marker, and *v ~ um* as the future tense marker. *u* is dropped when the set of suffixes beginning with vowel such as *um, a, aat, al* are added directly to the verb stem. *varu* changes into *va* before past tense marker and adverbial participle marker and *vaa* before imperative marker. *vaa (varu~vaa~va)* and *taa (~taa~ta)* are the two members of this class.

12. *akal*-class: The verbs of *akal* class take *nR* (i.e. *l+ndt>nR)* as the past tense marker and *kiR~kinR* as present tense marker, *v ~ um* as the future tense marker and *nRu* (*l+ndtu>nRu*) as the adverbial participle marker. The final *l* of the verbs becomes nil when it assimilates with the past tense suffix and adverbial suffix.

13. *kol*-class: The verbs of *kol* class take *nR* (l+*ndt>nR)* as the past tense marker, *kiR~kinR* as present tense marker, *v ~ um* as the future tense marker, and *nRu* (l+*ndtu>nRu)* as the adverbial participle marker. The final *l* becomes nil when it assimilates with the past tense suffix and adverbial suffix. The final *l* of the verbs geminates into *ll* when the set of suffixes beginning with vowel such as *um, a, aat, al* are added directly to the verbs.

14. *aaL*-class: The verbs of *aaL* class take *NT* (i.e. *L+ndt>NT)* as the past tense marker, *kiR~kinR* as present tense marker, *v ~ um* as the future tense marker, and *NTu* (i.e. *L+ndtu>NTu)* as the adverbial participle marker. The final *L* of the verbs becomes nil when it assimilates with the past tense suffix and adverbial suffix.

15. *koL*-class: The verbs of *koL* class take *ndt* (i.e. *L+ndt>NT) L* assimilates with the past tense suffix. *)* as the past tense marker, *kiR~kinR* as present tense marker, *v ~ um* as the future tense marker, and *NTu* (i.e. *L+ndtu>NTu)* as the adverbial participle marker. The final *L* becomes nil when it assimilates with the past tense suffix and adverbial suffix. The final *L* of the verbs geminates into *LL* when the set of suffixes beginning with vowel such as *um, a, aat, al* are added directly to the verbs.

16. *ooTu*-class: The verbs of *ooTu* class takes *in* as past tense marker, *kiR ~ kinR* as present tense marker, *v ~ um* as future tense marker and *i* as the adverbial participle marker. The final *u* of the verbs is deleted when the suffixes beginning with vowel such as *um, a, aat, al* are added to the verbs directly,

17. *col*-class: The irregular verb *col* is the only member of this group. The verb of *col*-class takes *nn* (*l+in> nn*) as past tense marker, *kiR ~ kinR* as present tense marker, *v ~ um* as future tense marker, and *i* (*in>i*) as the adverbial participle marker. The final *l* becomes nil when it assimilates with the past tense suffix. The final *l* of the verbs geminates into *ll* when the set suffixes beginning with vowel such as *um, a, aat, al, i* are added directly to the verbs.

18. *poo*-class: The verb of *poo*-class takes *n* (*in>n*) as the past tense marker, *kiR ~ kinR* as present tense marker, *v ~ um* as future tense marker, and y (*in>y*) as the adverbial participle marker. An increment *k* is added in between when the suffixes beginning with vowel such as *um, a, aat, al* are added to the verb stem directly. The verbs *poo* and *aa* are the two members of this class.

19. *kal*-class: The verbs of *kal* class take *R* (*l+t> RR*) as the past tense marker, *kiR~kinR* as present tense marker, *p ~ um* as the future tense marker, and *Ru* (*l+tu> RRu*) as the adverbial participle marker. The final *l* of the verbs changes into *R* when the suffix beginning with consonant such as the past tense suffix *R,* adverbial suffix *Ru,* present tense suffix *kiR ~ kinR,* future tense suffix *p* are added to the verbs directly. An increment *k* is added in between when the suffixes beginning with vowel such as *um, a, aata, al* are added to the verbs directly. The final *l* of the verbs changes into *R* even before the increment *k.*

20. *keeL*-class: The verbs of *keeL* class take *T* (*L+t>TT*) as the past tense marker and *kiR ~ kinR* as present tense marker, *p ~ um* as the future tense marker and *Tu* (*L+tu>TTu*) as the adverbial participle marker. The final *L* of the verbs changes into *T* when the suffixes beginning with consonant such as the past tense suffix *T,* adverbial suffix *Tu,* present tense suffix *kiR ~ kinR,* future tense suffix *p* are added to the verbs directly. An increment *k* is added in between when the suffixes beginning with vowel such as *um, a, aata, al* are added to the verbs directly. The final *L* of the verbs changes into *T* even before the increment *k.*

21. *uN*-class: The verbs of *uN*-class take *T* (i.e. *N+t>NT*) as the past tense marker, *kiR ~kinR* as present tense marker, *p ~ um* as the future tense marker and *Tu* as the

adverbial participle marker. The final *N* of the verbs geminates into *NN* when the suffixes beginning with vowel such as *um, a, aat, al* are added to the verb stem directly.

22. *puuN*-class: The verbs of *puuN*-class take *T* (i.e. *N+t>NT*) as the past tense marker and *kiR~kinR* as present tense marker and *p ~ um* as the future tense marker, and *NTu* (i.e. *N+tu>NTu*) as the adverbial participle marker.

23. *kaaN*-class: *kaaN* is the only member of this group. The verb of *kaaN*-class takes *T* (i.e. *N+t>NT*) as the past tense marker and *kiR~kinR* as present tense marker and *p ~ um* as the future tense marker and *NTu* (i.e. *N+tu>NTu*) as the adverbial participle marker.

24. *tin*-class: The verbs of *tin* class take *R* (i.e. *n+t>nR*) as the past tense marker and *kiR~kinR* as present tense marker, *p ~ um* as the future tense marker and *Ru* (i.e. *n+tu>nRu*) as the adverbial participle marker. The final *n* of the verbs geminates into *nn* when the suffixes beginning with vowel such as *um, a, aata, al* are added to the verbs directly.

25. *en*-class: The verbs of *en*-class take *R* (i.e. *n+t>nR*) as the past tense marker and *kiR~kinR* as present tense marker, *p ~ um* as the future tense marker and *Ru* (i.e. *n+tu>nRu*) as the adverbial participle marker. Unlike the previous class the final *n* of the verbs does not geminate into *nn* when the suffixes beginning with vowel such as *um, a, aata, al* are added to the verbs directly.

26. *ndil*-class: *ndil* is the only member of this class. The verb of *ndil* class take *nR (*i.e. *l+ndt>nR)* as the past tense marker, *kiR~kinR* as present tense marker, *p ~ um* as the future tense marker and *nRu* as the adverbial participle marker. The final *l* becomes nil when it assimilates with the past tense suffix *nR* and adverbial suffix *nRu*. The final *l* of the verbs changes into *R* when the suffixes beginning with consonant such as present tense suffix *kiR ~ kinR,* future tense suffix *p* are added to the verbs directly. An increment *k* is added in between when the suffixes beginning with vowel such as *um, a, aat, al* are added to the verbs directly. The final *l* of the verbs changes into *R* even before the increment *k.*

27. *paTi*-class: The verbs of *paTi* class take *tt* as the past tense marker, *kkiR~kkinR* as present tense marker, *pp ~ um* as the future tense marker and *ttu* as the adverbial participle marker. An increment *kk* is added in between when the suffixes beginning with vowel such as *um, a, aat, al* are added to the verb stem directly,

28. *ndaTa*-class: The verb of *ndaTa* class take *ndt* as the past tense marker and *kkiR ~ kkinR* as present tense marker, *pp ~ um* as the future tense marker and ndu as the adverbial participle marker. An increment *kk* is added in between when the suffixes beginning with vowel such as *um, a, aata, al* are added to the verb stem directly,

2.7. TAM chart

The flow chart given above depicts the analysis to be followed. Unlike the regular morphological analysis which is restricted within words and whose function to

strip off the affixes by assigning them meaning and identifying the lexeme to be matched with the item in the dictionary, the analysis of verbal complex goes beyond words as the verbal complex contains affixes as well as words which are stringed into a unit. So a TAM chart is prepared showing the morphemes (which includes suffixes as wells as auxiliary verbs) to be added to each verb at the horizontal axis and the verb classes at the vertical axis. (The TAM chart is divided into two parts to adjust with the space.)

Table 1: The vebal bases occurring at the first level and the suffixes occurring at the second level

Verb class	**Stem alternants**	**past tense suffix (occurring immediately after verb stem)**	**present tense suffix (occurring immediately after verb base)**	**future tense suffix (occurring immediately after verb base)**	**negative suffix (occurring immediately after verb base)**	**Infinitive suffix (occurring immediately after verb base)**	**Adverbial suffix (occurring immediately after verb base)**	**Nominizer 1 (occurring immediately after the verb)**	**Nominizer 2 (occurring immediately after the verb)**	**Nominer (occurring immediately after t verb)**
class 1 *cey* class	cey ~ ceyy	t	kiR ~ kinR	v ~ um	aa~aat	a	tu ~t	tal	al	kai
class 2 *azu* class	azu ~ az	t	kiR ~ kinR	v ~ um	aa~aat	a	tu~t	tal	al	kai
class 3 *caa* class	caa ~ ce	t	kiR ~ kinR	v ~ kum	kaa~kaat	ka	ttu~tt	tal	kal	kai
class 4 *viTu* class	viTu ~ viT	T	kiR ~ kinR	v ~ um	aa~aat	a	Tu~T	tal	al	kai
class 5 *peRu* -class	peRu~ peR	R	kiR ~ kinR	v ~ um	aa~aat	a	Ru~R	tal	al	kai
class 6 *pukaz* -class	pukaz	ndt	kiR ~ kinR	v ~ um	aa~aat	a	ndtu~ ndt	tal	al	kai
class 7 *vizu*- class,	vizu~ viz	ndt	kiR ~ kinR	v ~ um	aa~aat	a	ndtu~ ndt	tal	al	kai
class 8 *aRi*- class	aRi~ aRiy	ndt	kiR ~ kinR	v~um	aa~aat	a	ndtu~ ndt	tal	al	kai
class 9 *ndoo*- class	ndoo~ ndo~	ndt	kiR ~ kinR	v~kum	kaa~kaat	ka	ndtu~ ndt	tal	kal	kai
class 10 *vee*-	vee~ ve~	ndt	kiR ~ kinR	v~kum	kaa~kaat	ka	ndtu~ ndt	tal	kal	kai

class										
class 11 *vaa*-class,	vaa~ va~ varu~ var	ndt	kiR ~ kinR	v~um	aa~aat	a	ndtu~ ndt	tal	al	kai
class 12 *akal*-class	akal~ aka	nR	kiR ~ kinR	v~um	aa~aat	a	nRu~ nR	tal	al	kai
class 13 *kol*-class	kol ~ko ~ koll	nR	kiR ~ kinR	v~um	aa~aat	a	nRu~ nR	tal	al	kai
class 14 *aaL*-class	aaL~aa	NT	kiR ~ kinR	v~um	aa~aat	a	NTu ~NT	tal	al	kai
class 15 *koL*-class	koL~ko ~ koLL	NT	kiR ~ kinR	v~um	aa~aat	a	NTu~ NT	tal	al	kai
class 16 *ooTu* -class	ooTu~ ooT	in	kiR ~ kinR	v~um	aa~aat	a	i~iy	tal	al	kai
class 17 *col*-class	col~co~ coll	nn	kiR ~ kinR	v~um	aa~aat	a	i~iy	tal	al	kai
class 18 *poo*-class	poo	n	kiR ~ kinR	v~kum	kaa~kaat	ka	y	tal	kal	kai
class 19 *kal*-class	kal~ kaR	R	kiR ~ kinR	p~kum	kaa~kaat	ka	Ru~R	Ral	kal	kai
class 20 *keeL*-class	keeL~ keeT	T	kiR ~ kinR	p~kum	kaa~kaat	ka	Tu~T	tal	kal	kai
class 21 *puuN* -class	puuN	T	kiR ~ kinR	p~um	aa~aat	a	Tu~T	tal	al	kai
class 22 *uN*-calss	uN~uN N	T	kiR ~ kinR	p~um	aa~aat	a	Tu~T	tal	al	kai
class 23 *kaaN* -class	kaaN~ kaN	T	kiR ~ kinR	p~um	aa~aat	a	NTu	utal	al	kai
class 24 *tin*-class	tin ~tinn	R	kiR ~ kinR	p~um	aa~aat	a	nRu~ nR	utal	al	kai
class 25	en	R	kiR ~ kinR	p~um	aa~aat	a	nRu~ nR	utal	al	kai

***en-* class**										
class 26 *ndil-* class	ndil~ ndi~ ndiR ~ndill	nR	kiR ~ kinR	p~kum	kaa~kaat	ka	nRu	Ral	kal	kai
class 27 *paTi-* class	paTi	tt	kkiR ~ kkinR	pp~ kkum	kkaa~ kkaat	kka	ttu~ tt	ttal	kkal	kkai
class 28 *ndaTa-* class	ndaTa	ndt	kkiR ~ kkinR		kkaa~ kkaaat	kka	ndtu~ ndt	ttal	kkal	kkai

Table 2: The suffixes occurring at the third and fourth level

Verb class	**PNG suffix (occurring immediately after tense suffix)**	**Neuter PNG suffix (occurring immediately after the negative suffix *aa*)**	**Adverbial suffix (occurring immediately after negative *aa*)**	**Adjectival suffix (occurring immediately after tense suffix or negative suffix *aat*)**	**Conditional suffix (occurring immediately after past tense suffix)**	**Nominalizer 4 (occurring immediately after the negative *aa*)**	**Gerundival nominalizers occurring after adjectival participle)**	**Pronominalizers (occurring immediately after the adjectival participle)**
class 1 *cey* class	een/oom/aay /iir/iirkaL/ aan/ aaL/aar/ atu/ana	tu	mal~tu	a	aal	mai	tu	van/vaL/var/ varkaL/tu/vai
class 2 *azu* class	een/oom/aay /iir/iirkaL/ aan/ aaL/aar/ atu/ana	tu	mal~tu	a	aal	mai	tu	van/vaL/var/ varkaL/tu/vai
class 3 *caa* class	een/oom/aay /iir/iirkaL/ aan/ aaL/aar/ atu/ana	tu	mal~tu	a	aal	mai	tu	
class 4 *viTu* class	een/oom/aay /iir/iirkaL/ aan/ aaL/aar/ atu/ana	tu	mal~tu	a	aal	mai	tu	van/vaL/var/ varkaL/tu/vai
class 5 *peRu-* class	een/oom/aay /iir/iirkaL/ aan/ aaL/aar/ atu/ana	tu	mal~tu	a		mai	tu	van/vaL/var/ varkaL/tu/vai
class 6 *pukaz-*	een/oom/aay	tu	mal~tu	a	aal	mai	tu	van/vaL/var/ varkaL/tu/vai

class	/iir/iirkaL/ aan/ aaL/aar/ atu/ana							
class 7 *vizu*-class,	een/oom/a ay /iir/iirkaL/ aan/ aaL/aar/ atu/ana	tu	mal~tu	a	aal	mai	tu	van/vaL/var/ varkaL/tu/vai
class 8 *aRi*-class	een/oom/a ay /iir/iirkaL/ aan/ aaL/aar/ atu/ana	tu	mal~tu	a	aal	mai	tu	van/vaL/var/ varkaL/tu/vai
class 9 *ndoo*-class	een/oom/a ay /iir/iirkaL/ aan/ aaL/aar/ atu/ana	tu	mal~tu	a	aal	mai	tu	van/vaL/var/ varkaL/tu/vai
class 10 *vee*-class	een/oom/a ay /iir/iirkaL/ aan/ aaL/aar/ atu/ana	tu	mal~tu	a	aal	mai	tu	van/vaL/var/ varkaL/tu/vai
class 11 *vaa*-class,	een/oom/a ay /iir/iirkaL/ aan/ aaL/aar/ atu/ana	tu	mal~tu	a	aal	mai	tu	van/vaL/var/ varkaL/tu/vai
class 12 *akal*-class	een/oom/a ay /iir/iirkaL/ aan/ aaL/aar/ atu/ana	tu	mal~tu	a	aal	mai	tu	van/vaL/var/ varkaL/tu/vai
class 13 *kol*-class	een/oom/a ay /iir/iirkaL/ aan/ aaL/aar/ atu/ana	tu	mal~tu	a	aal	mai	tu	van/vaL/var/ varkaL/tu/vai
class 14 *aaL*-class	een/oom/a ay /iir/iirkaL/ aan/ aaL/aar/ atu/ana	tu	mal~tu	a	aal	mai	tu	van/vaL/var/ varkaL/tu/vai
class 15 *koL*-class	een/oom/a ay /iir/iirkaL/ aan/ aaL/aar/ atu/ana	tu	mal~tu	a	aal	mai	tu	van/vaL/var/ varkaL/tu/vai
class 16	een/oom/a ay	tu	mal~tu	a	aal	mai	tu	van/vaL/var/ varkaL/tu/vai

***ooTu*-class**	/iir/iirkaL/ aan/ aaL/aar/ atu/ana							
class 17 *col*-class	een/oom/a ay /iir/iirkaL/ aan/ aaL/aar/ atu/ana	tu	mal~tu	a	aal	mai	tu	van/vaL/var/ varkaL/tu/vai
class 18 *poo*-class	een/oom/a ay /iir/iirkaL/ aan/ aaL/aar/ atu/ana	tu	mal~tu	a	aal	mai	tu	van/vaL/var/ varkaL/tu/vai
class 19 *kal*-class	een/oom/a ay /iir/iirkaL/ aan/ aaL/aar/ atu/ana	tu	mal~tu	a	aal	mai	tu	van/vaL/var/ varkaL/tu/vai
class 20 *keeL*-class	een/oom/a ay /iir/iirkaL/ aan/ aaL/aar/ atu/ana	tu	mal~tu	a	aal	mai	tu	van/vaL/var/ varkaL/tu/vai
class 21 *puuN*-class	een/oom/a ay /iir/iirkaL/ aan/ aaL/aar/ atu/ana	tu	mal~tu	a	aal	mai	tu	van/vaL/var/ varkaL/tu/vai
class 22 *uN*-calss	een/oom/a ay /iir/iirkaL/ aan/ aaL/aar/ atu/ana	tu	mal~tu	a	aal	mai	tu	van/vaL/var/ varkaL/tu/vai
class 23 *kaaN*-class	een/oom/a ay /iir/iirkaL/ aan/ aaL/aar/ atu/ana	tu	mal~tu	a	aal	mai	tu	van/vaL/var/ varkaL/tu/vai
class 24 *tin*-class	een/oom/a ay /iir/iirkaL/ aan/ aaL/aar/ atu/ana een/oom/a ay /iir/iirkaL/ aan/ aaL/aar/ atu/ana	tu	mal~tu	a	aal	mai	tu	van/vaL/var/ varkaL/tu/vai
class 25 *en*-	een/oom/a ay	tu	mal~tu	a	aal	mai	tu	van/vaL/var/ varkaL/tu/vai

class	/iir/iirkaL/ aan/ aaL/aar/ atu/ana een/oom/a ay /iir/iirkaL/ aan/ aaL/aar/ atu/ana							
class 26 *ndil*- class	een/oom/a ay /iir/iirkaL/ aan/ aaL/aar/ atu/ana	tu	mal~tu	a	aal	mai	tu	van/vaL/var/ varkaL/tu/vai
class 27 *paTi*- class	een/oom/a ay /iir/iirkaL/ aan/ aaL/aar/ atu/ana een/oom/a ay /iir/iirkaL/ aan/ aaL/aar/ atu/ana	tu	mal~tu	a	aal	mai	tu	van/vaL/var/ varkaL/tu/vai
class 28 *ndaTa*- class	een/oom/a ay /iir/iirkaL/ aan/ aaL/aar/ atu/ana	tu	mal~tu	a	aal	mai	tu	van/vaL/var/ varkaL/tu/vai

Table 3: The third level auxiliary verbs occurring after infinitive and the suffixes of second occurring at the fourth level

Auxilary verbs occurring after infinitive form	**Past tense suffix (occurring immediately after verb stem)**	**Present tense suffix (occurring immediately after verb base)**	**Future tense suffix (occurring immediately after verb base)**	**Negative suffix (occurring immediately after verb base)**	**Infinitive suffix (occurring immediately after verb base)**	**Adverbial suffix (occurring immediately after verb base)**	**Nominizer 1 (occurring immediately after the verb)**	**Nominizer 2 (occurring immediately after the verb)**	**Nominizer 1 (occurring immediately after the verb)**
TTum *	-	-	-	-	-	-	-	-	-
muTi~ muTiy	ndt	kiR ~ kinR	um	aa~aat	a	ndtu~ ndt	-	-	-
iyal	-	-	um	aa~aat	a	-	-	-	-
kuuTum*	-	-	-	-	-	-	-	-	-
kuuT	-	-	um	aa~aat	-	i	-	-	-
paar	tt	kkiR ~ kkinR	pp~ kkum	kkaa~ kkaat	kka	ttu~ tt	ttal	kkal	kkai
iru	ndt	kkiR ~ kkinR	pp ~kkum	aa~aat	kka	ndtu~ndt	ttal	kkal	kkai

poo	n	kiR ~ kinR	v~kum	kaa~kaat	ka	y	tal	kal	kai
varu~var~ va~vaa (1)	ntd	kiR ~ kinR	v~um	aa~aat	a	nRu~ nR	tal	al	kai
varu~var~ va~vaa (2)	-	kiR~kinR	um	aa~aat	a	ndtu~ ndt	-	-	-
veeNT	-	-	um	-	-	i ~iy	-	-	-
veeNTivaru ~ va	ndt	kiR~kinR	um	aa~aat	a	ndtu ~ndu	tal	al	kai
maaTT	-	-	-	-	-	-	-	-	-
illai *	-	-	-	-	-	-	-	-	-
cey~ceyy	t	kiR ~ kinR	v ~um	aa~aat	a	tu~t	tal	al	kai
vai	tt	kkiR ~ kinR	pp ~ kkum	kkaa ~ kkaat	a	ttu ~ t	ttal	kkal	kkai
paTu ~ paT	T	kiR ~ kinR	v ~ um	aa ~ aat	a	Tu ~ T	tal	al	kai

Table 4: The third level auxiliary verbs occurring after infinitive and the suffixes of third level occurring at the fifth level and the suffixes of fourth level occurring at the sixth level

Auxilary verbs occurring after infinitive form	**PNG suffix (occurring immediately after tense suffix)**	**Neuter PNG suffix (occurring immediately after the negative suffix *aa*)**	**Adverbial suffix (occurring immediately after negative *aa*)**	**Adjectival suffix (occurring immediately after tense suffix or negative suffix *aat*)**	**Conditional suffix (occurring immediately after past tense suffix)**	**Nominalizer 4 (occurring immediately after the negative *aa*)**	**Gerundival nominalizers occurring after adjectival participle)**	**Pronominalizers (occurring immediately after the adjectival participle)**
TTum *	-	-	-	-	-	-	-	-
muTi~ muTiy	-	tu	mal~tu	a	aal	mai	tu	van/vaL/var/ varkaL/tu/vai
iyal	-	tu	mal~tu	-	-	mai	tu	van/vaL/var/ varkaL/tu/va
kuuT		tu	-	a	-	-		van/vaL/var/ varkaL/tu/vai
paar	een/oom/aay /iir/iirkaL/ aan/ aaL/aar/ atu/ana	tu	mal~tu	a	aal	mai	tu	van/vaL/var/ varkaL/tu/vai
iru	een/oom/aay /iir/iirkaL/ aan/ aaL/aar/ atu/ana	tu	mal~tu	a	aal		tu	van/vaL/var/ varkaL/tu/vai

poo	een/oom/aay /iir/iirkaL/aan/ aaL/aar/ atu/ana	tu	mal~tu	a	aal	mai	tu	van/vaL/var / varkaL/tu/vai
varu~var~va~vaa (1)	een/oom/aay /iir/iirkaL/aan/ aaL/aar/ atu/ana	tu	mal~tu	a	aal	mai	tu/vai	van/vaL/var / varkaL/tu/vai
varu~var~va~vaa (2)	een/oom/aay /iir/iirkaL/aan/ aaL/aar/ atu/ana	tu	mal~tu	a	aal	mai	tu	van/vaL/var / varkaL/tu/vai
veeNT-	-	-	-	a	-	-	tu	van/vaL/var / varkaL/tu/vai
maaTT	een/oom/aay /iir/iirkaL/aan/ aaL/aar/ atu/ana	tu	-	-	-	-	tu	van/vaL/var / varkaL/tu/vai
illai *	-	-	-	-	-	-	-	-
cey~ceyy	een/oom/aay /iir/iirkaL/aan/ aaL/aar/ atu/ana	tu	mal~tu	a	aal	mai	tu	van/vaL/var / varkaL/tu/vai
vai	een/oom/aay /iir/iirkaL/aan/ aaL/aar/ atu/ana	tu	mal~tu	a	aal	mai	tu	

Table 5: The third level auxiliary verbs occurring after adverbial participle and the suffixes of second level occurring at the fourth level

The auxiliary verbs occurring after adverbial participle	**past tense suffix (occurring immediately after verb stem)**	**present tense suffix (occurring immediately after verb base)**	**future tense suffix (occurring immediately after verb base)**	**negative suffix (occurring immediately after verb base)**	**Infinitive suffix (occurring immediately after verb base)**	**Adverbial suffix (occurring immediately after verb base)**	**Nominizer 1 (occurring immediately after the verb)**	**Nominizer 2 (occurring immediately after the verb)**	**Nominizer 3 (occurring immediately after the verb)**
aayiRRu *									
iru (class	ndt	kkiR~kkinR	pp~kkum	kkaa	a	ndtu ~ndt	ttal	al	kkai

viTu	T	kiR~kinR	v~um	aa	a	Tu~T	tal	al	kai
poo	n	kiR~kinR	v~kuu	kaa	a	y	tal	kal	kai
tiir	tt	kiR~kiR	pp~kkum	aa	ka	ttu~tt	ttal	kkal	kkai
muTi	tt	kkiR	pp~kkum	kkaa	kka	ttu~tt	ttal	kkal	kkai
pooTu	T	kiR	v~um	aa	a	Tu~T	tal	al	kai
koNTiru	ndt	kkiR	pp~kkum	kkaa	kka		tta	kkal	kkai
koL	NT	kiR	v~	aa	a		tal	al	kai
paar	tt	kkiR~kkinR	pp~kum	kkaa	kka	ttu~tt	ttal	kkal	kkai
koTu	tt	kkiR~kinR	pp~kum	kkaa	kka	ttu~tt	ttal	kkal	kkai
tolai	tt	kkiR~kinR	pp~kum	kkaa	kka	ttu~tt	ttal	kkal	kkai
kizi	tt	kkiR~kinR	pp~kum	kkaa	kka	ttu~tt	ttal	kkal	kkai
taLLu	in	kiR~kin	v~um	aa	a	i	tal	al	kai

Table 6: Table 5: The third level auxiliary verbs occurring after adverbial participle and the suffixes of third level occurring at the fifth level and the suffixes of fourth level occurring at the sixth level

The auxiliary verbs occuring after adverbial participle	**PNG suffix (occurring immediately after tense suffix)**	**Neuter PNG suffix (occurring immediately after the negative suffix *aa*)**	**Adverbial suffix (occurring immediately after negative *aa*)**	**Adjectival suffix (occurring immediately after tense suffix or negative suffix *aat*)**	**Conditional suffix (occurring immediately after past tense suffix)**	**Nominalizer 4 (occurring immediately after the negative *aa*)**	**Gerundival nominalizers occurring after adjectival participle)**	**Pronominalizers (occurring immediately after the adjectival participle)**
aayiRRu *	-	-	-	-	-	-	-	-
iru	een/oom/aay /iir/iirkaL/ aan/ aaL/aar/ atu/ana	tu	mal~tu	a	aal		tu	van/vaL/var/ varkaL/tu/vai
viTu	een/oom/aay /iir/iirkaL/ aan/ aaL/aar/ atu/ana	tu	mal~tu	a		mai	tu	
poo	een/oom/aay /iir/iirkaL/ aan/ aaL/aar/ atu/ana	tu	mal~tu	a	aal	mai	tu	van/vaL/var/ varkaL/tu/vai
tiir	een/oom/aay /iir/iirkaL/ aan/ aaL/aar/ atu/ana	tu	mal~tu	a		mai	tu/vai	van/vaL/var/ varkaL/tu/vai
muTi	een/oom/aay /iir/iirkaL/	tu	mal~tu	a	aal	mai	tu	van/vaL/var/ varkaL/tu/v

	aan/ aaL/aar/ atu/ana							ai
pooTu	een/oom/a ay /iir/iirkaL/ aan/ aaL/aar/ atu/ana	tu	mal~tu	a	aal	mai	tu	van/vaL/var / varkaL/tu/v ai
koNTiru	een/oom/a ay /iir/iirkaL/ aan/ aaL/aar/ atu/ana	tu	mal~tu	a	aal	mai	tu	van/vaL/var / varkaL/tu/v ai
koL	een/oom/a ay /iir/iirkaL/ aan/ aaL/aar/ atu/ana	tu	mal~tu	a	aal	mai	tu	van/vaL/var / varkaL/tu/v ai
paar	een/oom/a ay /iir/iirkaL/ aan/ aaL/aar/ atu/ana	tu	mal~tu	a	aal	mai	tu	van/vaL/var / varkaL/tu/v ai
koTu	een/oom/a ay /iir/iirkaL/ aan/ aaL/aar/ atu/ana	tu	mal~tu	a	aal	mai	tu	van/vaL/var / varkaL/tu/v ai
tolai	een/oom/a ay /iir/iirkaL/ aan/ aaL/aar/ atu/ana	tu	mal~tu	a	aal	mai	tu	van/vaL/var / varkaL/tu/v ai
kizi	een/oom/a ay /iir/iirkaL/ aan/ aaL/aar/ atu/ana	tu	mal~tu	a	aal	mai	tu	van/vaL/var / varkaL/tu/v ai
taLLu	een/oom/a ay /iir/iirkaL/ aan/ aaL/aar/ atu/ana	tu	mal~tu	a	aal	mai	tu	van/vaL/var / varkaL/tu/v ai

Table 7: The auxiliary verbs occuring after *al*-nominalized form

Verbs	**past tense suffix (occur**	**present tense suffix (occurri**	**future tense suffix (occurrin**	**negative suffix (occurrin g**	**Infinitive suffix (occurrin g**	**Adverbi al suffix (occurri ng**	**Nominiz er 1 (occurri ng**	**Nominiz er 3 (occurri ng**	**Nomin izer 3 (occur ring**

	ring immediately after verb stem)	**ng immediately after verb base)**	**g immediately after verb base)**	**immediately after verb base)**	**immediately after verb base)**	**immediately after verb base)**	**immediately after the verb)**	**immediately after the verb)**	**immediately after the verb)**
aam#									
aakum#									
aakaatu#									

Table 8: The auxiliary verb occuring after *atu*-nominalized firm

Verbs				
uNTu#				
illai#				
kiTaiyaatu#				

(The marker # in the above tables denotes that the verbal forms are the finite forms and cannot be inflected further.)

The programmes operate on the TAM chart and give the analytical and synthetic output.

Finite state automata and morphological analysis

Computational analysis of verbal complex involves two processes:

1. The analysis of verbal complex into minimal meaningful units (which includes suffixes and bound auxiliary verbs) and assigning semantic features to each unit and

3. The generation of verbal complex out of the same meaningful units.

Let us make an assumption that the building of verbal complex is the concatenation of morphemes and bound words. Clearly the simplest way to build such a model would be to assume that the allowability of a particular morpheme or a bound word in a given context depends only upon the morpheme that precedes it. Take the word *ceytuviTTeen*, whose morphological structure is *cey+tu + viT+T+een*. The combination of *cey* which is a verb and *tu* which an adverbial marker forms the first chunk of the verbal complex and the combination of *viT* which is a secondary verb, *T* which is a past tense marker and *een* which is a person-number-marker forms the second chunk of the word. Both of the chunks (V1 and V2) make a verbal complex.

Finites state Automata

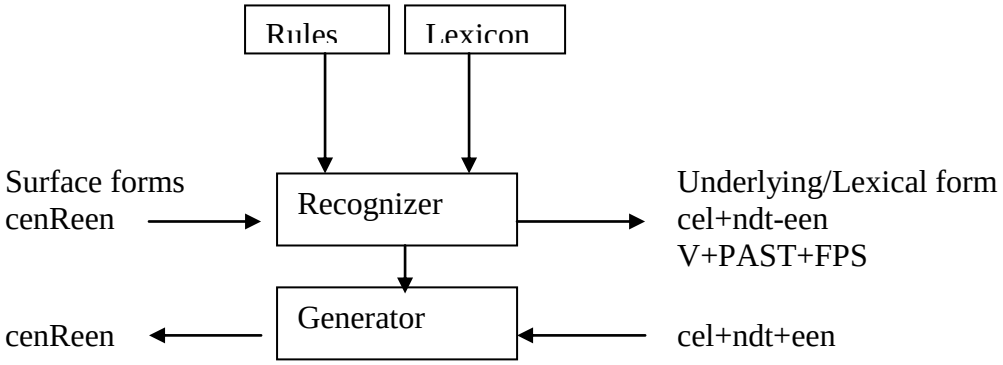
One can easily model such a system with a finite-state machine (finite-state automaton or finite-state transition network) of the kind familiar from formal language theory. More formally, a finite automation is defined by the following five parameters:

Q: a finite set of N states q0, q1...,qN
Σ: a finite input alphabet of symbols
q0: the start state
F: the set of final states, $F \subseteq Q$
δ (q,i): the transition function or transition matrix between states. Given a state $q \in Q$ and an input symbol $i \in \Sigma$, δ (q,i) returns a new state $q' \in Q$. δ is thus a relation from $Q \times \Sigma$ to Q;

A set of rewrite rules, or grammar, can be thought of not only as an intentional definition of the units of a language, but also as the description of an abstract machine which is able to accept (or generate) just those strings which belong to the language. These machines are called (abstract) automata, and the complexity of their potential behaviour is directly determined by the internal complexity of the rules of the grammar to which they correspond. All such machines may be characterized in terms of a set of states, an input device which can access one input symbol at a time, and control unit which can examine and read input and cause the machine to shift from one sate to another. In addition, such machines may have memory, which can be accessed by the control unit for storing and testing symbols. The computational power of the machine is determined essentially by the complexity of its memory and of the operations that can be performed on the memory. The simplest of these automata are called finite state machines or finite state automata (FSA), and the languages they can accept are regular languages.

PCKimmmo

The two level morphology of Kosekenniemi has been a widely implemented one, particularly in European languages. PCKimmo can do two functions as a morphological processor, recognition. In the case of the generator component of PCKimmo, it takes a lexical form of the foot as input. To this imput rules recorded in the rules file are applied and the corresponding surface surface form is returned. It does not use the lexicon. The recognizer accepts as input a surface form, applies the rules, consults the lexicon and returns the corresponding lexical form and gloss string. What makes it easier for the user and important from the point of economy of model is the fact that the same set of rules can be used for both the functions of recognition and gneration. The rules can be used birectionally (1990:8). If certain phonological, or of for that matter a morphological or orthographic, rule is written to refer to a transformation form the underlying structure to surface structure, they need not be rewritten for the reverse process. In reverse, it processes form the surface form to the underlying form. This model can be represented through the following diagram (Simons 1989).



(FPS=First person singular)

The two levels used in this model are linguistic descriptions that can handle any tow levels of abstractions, be it allophonic alternations, morphophonemic alternations or orthographic variations.

API Morphological Analyzer

The API Morphological analyzer developed by AUKBC Research Centre in collaboratin with the first author of the paper depends on automata table present in its data file, statetable.dat. The API morphological Analyzer is actually designed to strip a word of its inflections.

The machine while processing is set to be at the state beging represented by 0:0. Then the input is looked into root dictionary for a match. If a match is found it returns it along with its gloss and by printing, 'FINISH' in the command line. If it doesn't find the input in the root file, then it starts matching characters form the right preriphery with the entries listed in the file 'statetable.dat.' If it doesn't find a match there then it goes and checks in a file called 'SANDHI.dat'. If it finds a morpheme, then it strips the word of the morpheme, takes the remaining part of the word to be checked again in a file called 'ROOT.dat'. The sandhi replacement tables are in the file 'sandhidict.dat'. For all these operations the analyzer depends on automata tables present in the datafile 'statetable.dat'. This table contains various rules defining the paths needed to identify the affixes in a Tamil word. The state table is based on the verb flow chart. The format in which the rules are represented in this file is as fallows:

Stateno:morpheme, next state [-nextstate...]; [morpheme, nextstate];

An example for a morpheme in the suffix file will be as follows:

Ex.3.0:avan,1.1;avaL,1.1; avar,1.1;atu,1.1;avai,1.1; ana,1.1; kai,2.1.;tal,2.1; #ttal,2.1-End;Ral,2.1.-End

The fist entry in the line (stateno) is an arbitray number (or word) designating the state that the automata must be in when it scans the line. This 'stateno' is separated form other entries by a colon. Any number of entries may follow the 'stateno'. A semicolon terminates each entry. The semicolon is only optional for the last entry.

Each entry of two fields, the morpheme and the number of the next state that the automata will transit into if the morpheme in the entry is present at the end of word. These two fields are separated by a coma.

The following is an example for a rule in the file which strips the word the change because of the sandhi operations.

Ex. maram+ai==> marattai

The correspondence shows all that *m* will be replaced by *tt* when *ai* is added. During analysis tt will be replaced by the original *m* deletion. The entry in the rule file will be in the following format.

Ex.
4a +/-
<replacementCharacter>, {{Suffix{+postfix:[postfix...]};
{Suffix{+postfix:[postfix...]};

The morpheme entry will be as follows:

Ex.4b. –m, {tt{+ai:+aal:+uTan:+in:+ttin:+il:}; ng{+kaL:}}

The '-' sign before *m* indicates that the character will be added. The suffixes are enclosed in curly brackets. In the case of multiple suffixes, as is the case above, semicolons separate each suffix.

Different modules are used to do different functions. The modules have been written in the compiled languge Java. The main functions load is done by TamilMorph.java. A module called RootDict. java does the search for the match in the root file. Similarly Suffix.java, Sandhi.java etc. take care of various functions using the finite state automata.

4. Conclusion

The preparation of morphological analyzer is the preliminary step to be made before venturing into any other type of natural language processing. Tamil being a complex inflectional language is a challenge for computational linguists who try to prepare a morphological analyzer for it. There are attempts to prepare morphological generators and analyzers for Tamil. Preparing a morphological generator or analyzer for the purpose of demonstration is quite different from preparing an exhaustive and efficient working system. It is not too difficult to prepare a system with the coverage of 60 to 65%. However increasing the efficiency beyond 65% and to reach a realistic level of 95% to 97% is a hard task. Most of the existing descriptions focus on a narrow and vertical detail rather than on breadth and exhaustiveness. Generally the coverage obtained by incorporating the descriptions of the existing work is never beyond 50%. The evaluation method points out that the efficiency of the morphological analyzer prepared for Tamil by the present team is very remarkable.

The preparation of a morphological analyzer, in which the processing of verbal complex forms a part, has may natural language applications such as parsing,

text generation, machine translation, preparing dictionary tools and lemmatization. It also helps in speech applications such as text-to-speech synthesizing and speech recognition and in word processing applications like spell checking and text input and in retrieval of documents. Finding the categorical details of the verbal forms helps in assigning parts-of-speech taggs to the verbal complex.

Acknowledgement

My thanks are due to Dr.S.Boopathy, who was the chief investigator for the UGC sponsored porject entitled "Spell and grammar checker for Tamil"and Dr. Arulmozi who was working as the Research Assistant in the same project. (I was working in the project as co-investigator.) Our thanks are due to the members of the NLP team of AUKBC Research Centre, MIT Campus Chennai, Mr. S.Visvanathan, and Mr. S. Ramesh Kumar who are involved with me in the preparation of morphological analyzer for Tamil.

REFERENCE

Agesthialingom, S. 1964. 'Auxiliary verbs in Tamil'. Tamil culture 11:3.

Annamalai,E. 1985. Dynamics of verbal extension. Trivandrum: Dravidian Linguistics Association.

Dale, I.R.H. 1980. 'Syntax, Semantics, and Tamil auxiliary verbs'. In: Auxiliaries in Dravidian. (eds.). S.Agestialingom & G. Srinivasavarma. Annamalainagar: Annamalai University.

Hook, P.E. 1974. The compound verb in Hindi. Ann Arbor: University of Michigan.

Joseph, N. 1983. Auxiliaries in Tamil. Ph.D Thesis. Annamalainagar: Annamalai University.

Jurafsky, D and Matin, J.H. 2000. Speech and language processing: an introduction to natural language processing, computational linguistics, and speech recognition. New Jersey: Prentice Hall.

Lehman, T. 1993.A grammar of Modern Tamil. Pondicherry: Pondicherry Institute of Linguistics and Culture.

Rajendran, S. 1999. 'Spell and grammar checker for Tamil'. Paper read in 27th All India Conference of Dravidian Linguists held in ISDL, Thiruvananthapuram, 17th-19th, July, 1999.

----------. 2001. Wordfomation in Tamil (manuscript). Thanjavur: Tamil University.

Rajendran, S., Arulmozi, S., Ramesh Kumar, Viswanathan, S. 2001. 'Computational morphology of verbal complex'. Paper read in Conference at Dravidan University, Kuppam, December 26-29, 2001.

Rangan, K. 1970. 'Modals as main verbs in Tamil'. In Proceedings of the First All India Conference of Linguists. Poona: Deccan College.

Ranganathan, Vasu. 1997. 'A Lexical Phonological Approach to Tamil Word by computer'. International Journal of Dravidian Linguistics 26.1:57-70.

Rao, Umamaheshwar G. 1996. 'Compound verb formation in Telugu'. Paper presented in the National Workshop-cum-seminar on Lexical Typololgy, Telugu University, Hyderabad

Roper, T. and Siegel, D. 1978. 'A lexical transformation for verbal compounds'. Linguistic Inquiry 9.199-260.

Sengupta, G. 1997. 'Three Models of Morphological Processing.' South Asian Language Review, 8:1, 1-26.

----------. 1998. 'Language Technology and Morphological Processing'. The Administrator, 43:3.

Simons, G. 1989. 'A tool for exploring morphological' Notes on Linguistics 44:51-59.

Singh, J. 1994. 'Compound verb in Hindi: A GB approach'. Paper presented in National seminar on Word formation in Indian Languages, Hyderabad, Osmania University, February 9-10, 1994.

Sproat, R. 1992. Morphology and computation. Cambridge: The MIT Press.

Steever, S.B. 1983. A study in auxiliation: the grammar of the indicative auxiliary verb system of Tamil. Unpublished Ph.D. dissertation. Chicago: University of Chicago.

----------. 1993. Analysis to synthesis: the development of complex verb morphology in Dravidian languages. New York: Oxford University Press.